IRE Transactions

on INFORMATION THEORY

VOLUME IT-1 DECEMBER 1955 NUMBER 3

Published Quarterly

PUBLISHED BY THE
Professional Group on Information Theory

IRE PROFESSIONAL GROUP ON INFORMATION THEORY

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of the prescribed annual assessment of $2.00.

ADMINISTRATIVE COMMITTEE

IRE TRANSACTIONS

on Information Theory

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Information Theory at 1 East 79th Street, New York 21, N. Y. Responsibility for the contents rests upon the authors, and not upon the Institute, the Group or its members. Single copy prices: IRE-PGIT members, $1.55; IRE members, $2.30; nonmembers, $4.65.

L. A. DE ROSA

L. A. de Rosa was born in Tenafly, N. J., on September 11, 1910. He received the BS degree in electrical engineering from the Polytechnic Institute of Brooklyn, where he continued with graduate work until 1934.

Mr. de Rosa's professional work began in 1931, as an engineer with the DeForest Tube Company, Newark, N. J. In 1932, he became identified with Electrad, Inc., New York City, as a research engineer on radio components, receivers, and amplifiers. From 1934 to 1937, Mr. de Rosa was chief engineer of Electrotechnical Laboratories, New York City, where he engaged in work on electronics, remote control, and recording. In 1937, he became chief engineer with the Electrad Division of P. R. Mallory Co., Indianapolis, Ind., where he worked for a year on photoelastic research and special devices. The following year, while at Indianapolis, Mr. de Rosa devoted individual research to physio-psychological acoustics. From 1939 until 1942, he occupied a position as staff engineer with the Electronics Research Laboratories of the National Cash Register Co., Dayton, Ohio. Here his work was concerned with magnetic-materials research, electronic computers, and pulse communication.

From 1942 until the present time, Mr. de Rosa has been with the Federal Telecommunication Laboratories, Nutley, N. J. First as a senior project engineer, and later as a department head, he was engaged in research work on aerial navigation, direction finders, and automatic-landing equipment. From 1945 until 1953, he served as a division head in charge of acoustical research and electronic countermeasures work. Since 1953, Mr. de Rosa has been head of the Electronic Countermeasures Laboratory, which now comprises a group of departments concerned with research and development in the fields of active and passive electronic countermeasures systems, antenna research, data processing, and communication theory.

Mr. de Rosa has presented various technical papers at conventions and meetings of the Institute of Radio Engineers, the Radio Club of America, and the Acoustical Society of America. He holds about forty patents issued in the fields of direction finding, radar, antennas, communications, and components.

A Fellow of the Institute of Radio Engineers and chairman of its Professional Group on Information Theory, Mr. de Rosa also holds membership in the American Physical Society and the Acoustical Society of America.

In Which Fields Do We Graze?

L. A. DE ROSA

Chairman, Professional Group on Information Theory

THE EXPANSION of the applications of Information Theory to fields other than radio and wired communications has been so rapid that oftentimes the bounds within which the Professional Group interests lie are questioned. Should an attempt be made to extend our interests to such fields as management, biology, psychology, and linguistic theory, or should the concentration be strictly in the direction of communication by radio or wire?

To make one's interest the formulation and extension of the general theory of information, and then, having armed oneself with such a universal and powerful tool, to consider only those applications which deal with radio and wire communication, is an attitude which has been criticized by a number of our members.

Other Professional Groups whose interests lie in more sharply defined fields must, perforce, consider the application of Information Theory to their respective fields; otherwise, the benefits which may accrue through the extension of Information Theory to these various specialized fields might occur belatedly, or not at all.

Some of our members argue that should the application of Information Theory to other specialized fields be left to their specialists and the interests of PGIT not extend to fields other than radio and wire communication, then PGIT would be a purely academic and theoretical group with no interests in any but the general, universally applicable, mathematical procedures.

We have heard the opposite views expressed also, namely that PGIT should encourage the extension of the theorems to other general fields and broaden the scope of PGIT to include the interests of Psychology, Biology, and other branches of the "Arts and Sciences." In so doing, it is argued, PGIT becomes a creative group in advancing the theory of information and in assisting other Professional Groups. Thus, by disseminating information of other fields which may be required for the over-all solution of the problem of communication from one subjective sensory terminal to another (the over-all "brain-to-brain" terminals), a *raison d'être* is established for us.

At least one more group feels that PGIT should confine itself to adapting the generic developments of Information Theory to the specific field of radio, electronics, and wire communication, foregoing all ties with computers, television, telemetry, management, automation, or circuit theory.

It would be interesting to obtain the views of PGIT members with regard to the proper bounds of our interests and activities, for without such expression, proper direction cannot be achieved.

Theory of Noise in a Correlation Detector*

M. HOROWITZ AND A. A. JOHNSON†

Summary—The problem of detecting a signal that has both magnitude and sign is considered. A new type of correlation device, which employs the derivative of the correlation function, is proposed for measuring target range and position, and an analysis to determine the most useful waveform is made. The effects of white noise accompanying the input signal are minimized when an input waveform of long duration, wide bandwidth, and high derivative power is chosen.

Introduction

It HAS BEEN recognized for some time that correlation detectors may be used to detect signals contaminated by noise. Fano[1] has studied the particular case of a correlation detector consisting of a multiplier followed by a low-pass filter. Goldman[2] has summarized the work of Lee[3] in using cross-correlation for detecting repetitive signals. Woodward,[4] in his simple theory of radar reception, employs the correlation function.

It is the purpose of this paper to extend the abovementioned earlier work in order to deal with the problem of detecting a vector signal; *i.e.*, one that has both magnitude and direction. In particular, whereas Woodward has considered the problem of measuring the scalar range to a target by analysis of transmitted and received waveforms, we shall be concerned not only with measuring the range to the target but also with finding out whether the target is in front or to the rear of the receiver's position.

A Special Type of Correlation Detector

We consider a real wave form $f(t)$ where f is amplitude and t is time. The function $f(t)$ is assumed to be a stationary time series; in this case, as Wiener[5] has remarked, $f(t)$ possesses an autocorrelation function

$$\phi(\tau) \equiv \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)f(t+\tau)\,dt.$$

Since $f(t)$ is assumed real, $\phi(\tau)$ is a real symmetric function of τ, as is noted by Bartlett.[6] The first derivative of $\phi(\tau)$ with respect to τ, $d\phi/d\tau$, will hence be an odd function; this is shown, for example, by Indjoudjian.[7]

* Received by PGIT October 10, 1955.
† Aerophysics Depts., Goodyear Aircraft Corp., Akron, Ohio.
1 R. M. Fano, "Signal-to-noise ratio in correlation detectors," *M.I.T. Res. Lab. Elec. Tech. Rep.*, No. 186; February, 1951.
2 S. Goldman, "Information Theory," Prentice-Hall, Inc., New York, N. Y., p. 279, 1953.
3 Y. W. Lee, "Application of statistical methods to communication problems," *M.I.T. Res. Elec. Tech. Rep.*, No. 181; 1950.
4 P. M. Woodward, "Probability and Information Theory, with Applications to Radar," McGraw-Hill Book Co., Inc., New York, N. Y., p. 82, 1953.
5 N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y., p. 18; 1949.
6 M. S. Bartlett, "An Introduction to Stochastic Processes," Cambridge University Press, Cambridge, England, p. 160; 1955.
7 L. de Broglie et al., "La Cybernetique," Editions de la Revue d'Optique, Paris, France, p. 49; 1951.

We now define a special type of correlation detector that computes a time shift τ by means of the function

$$\alpha(\tau) = \frac{1}{T}\int_0^T f'(t)f(t+\tau)\,dt, \tag{1}$$

where T is the averaging time, τ is a time shift whose magnitude and sign are to be determined, and prime denotes differentiation with respect to t.

It is assumed that $f(t)$ possesses finite derivatives so that $f(t+\tau)$ may be expanded in a Taylor series; it is assumed the series converges. In this case, we find, from (1), that

$$\alpha(\tau) = \frac{1}{T}\frac{[f(T)]^2 - [f(0)]^2}{2} + \tau\left\{\frac{1}{T}\int_0^T [f'(t)]^2\,dt\right\} + \cdots.$$

For sufficiently large T and small τ we have approximately

$$\alpha(\tau) = \tau\left\{\frac{1}{T}\int_0^T [f'(t)]^2\,dt\right\}. \tag{2}$$

$\alpha(\tau)$ is thus essentially an odd function for small τ and large T and has the form shown in Fig. 1. The odd

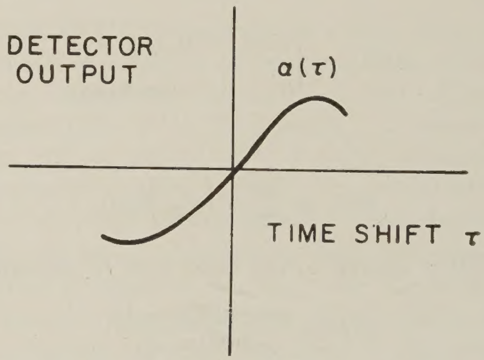


Fig. 1
Form of output function.

character of $\alpha(\tau)$ follows also from the fact that $d\phi/d\tau$ is odd and that $\alpha(\tau)$ will not differ much from $d\phi/d\tau$ when T is large and τ is small. Then by (1) and (2), for small τ and sufficiently large T,

$$\tau \approx \frac{1}{TP'}\int_0^T f'(t)f(t+\tau)\,dt, \tag{3}$$

where P' is the average power in $f'(t)$; that is to say,

$$P' \equiv \frac{1}{T}\int_0^T [f'(t)]^2\,dt.$$

We therefore define the output signal in the noise-free case as

$$s(\tau) = \frac{1}{TP'}\int_0^T f'(t)f(t+\tau)\,dt. \tag{4}$$

When the waveform $f(t)$ is accompanied by input noise $n(t)$, the output of the detector will be $s(\tau)$ plus output noise. The object of our analysis is to find a suitable wave form $f(t)$ so that the signal $s(\tau)$ can be most readily extracted from the output of the detector.

Evaluation of Output Signal-to-Noise Power

Let us now assume that a noise $n(t)$ accompanies the function $f(t + \tau)$ in (4) so that the input to the correlation detector is

$$g(t + \tau) \equiv f(t + \tau) + n(t)$$

instead of $f(t + \tau)$. Substituting the above expression for $g(t + \tau)$ in the place of the $f(t + \tau)$ in the right-hand side of (4), we find that the output of the correlation detector takes the form

$$\frac{1}{TP'} \int_0^T f'(t)f(t + \tau)\, dt + \frac{1}{TP'} \int_0^T f'(t)n(t)\, dt. \tag{5}$$

Hence, the output noise is given by

$$N_o = \frac{1}{TP'} \int_0^T f'(t)n(t)\, dt. \tag{6}$$

We now seek a suitable representation of $f'(t)$ and $n(t)$. We have chosen $f'(t)$ and $n(t)$ as functions with band-limited spectra to take advantage of the properties of the Whittaker cardinal functions.[8–10] Furthermore, we define

$$f_T'(t) = \begin{cases} f'(t) & 0 \leqq t \leqq T \\ 0 & \text{elsewhere.} \end{cases}$$

Then

$$f_T'(t) \approx \sum_{k=1}^{2TW} f'\left(\frac{k}{2W}\right)u_k(t), \tag{7}$$

where $f'(t)$ is limited to the band 0 to W cps and

$$u_k(t) = \frac{\sin \pi(2Wt - k)}{\pi(2Wt - k)}.$$

We now consider white noise as described by Shannon,[11]

$$n(t) = \sum_{k=-\infty}^{\infty} n\left(\frac{k}{2W_n}\right) \frac{\sin \pi(2W_n t - k)}{\pi(2W_n t - k)},$$

with the $n(k/2W_n)$ normal and independent all with the same standard deviation $\sqrt{N}$. This equation is a representation of white noise, band-limited to the band 0 to W_n cycles per second and with average power N. We now assume $W_n = W$. This assumption is possible since in either case we may choose the greater limit. Then

$$n(t) = \sum_{k=-\infty}^{\infty} n\left(\frac{k}{2W}\right)u_k(t). \tag{8}$$

Hence,

$$N_o = \frac{1}{TP'} \int_{-\infty}^{\infty} \left[\sum_{k=1}^{2WT} f'\left(\frac{k}{2W}\right)u_k(t)\right]\left[\sum_{k=1}^{2WT} n\left(\frac{k}{2W}\right)u_k(t)\right] dt. \tag{9}$$

By the orthonormality relation,

$$\int_{-\infty}^{\infty} u_k u_l\, dt = \delta_{kl} \frac{1}{2W},$$

we have

$$N_o = \frac{1}{2WTP'} \sum_{k=1}^{2WT} f'\left(\frac{k}{2W}\right)n\left(\frac{k}{2W}\right). \tag{10}$$

We wish to know the ensemble average $\langle N_o^2\rangle_{AV}$ when $f(t)$ is kept fixed and $n(t)$ varies over the ensemble of noise functions. We have

$$\langle N_o^2\rangle_{AV} = \frac{1}{(2WTP')^2} \left\langle \left[\sum_{k=1}^{2WT} f'\left(\frac{k}{2W}\right)n\left(\frac{k}{2W}\right)\right] \times \cdot \left[\sum_{k=1}^{2WT} f'\left(\frac{k}{2W}\right)n\left(\frac{k}{2W}\right)\right]\right\rangle_{AV}. \tag{11}$$

Because the noise is white and Gaussian,

$$\left\langle n\left(\frac{k}{2W}\right)n\left(\frac{l}{2W}\right)\right\rangle_{AV} = N\delta_{kl}. \tag{12}$$

The integral defined above,

$$P' \equiv \frac{1}{T} \int_0^T [f'(t)]^2\, dt,$$

may be approximated by the series

$$\frac{1}{2TW} \sum_{k=1}^{2TW} \left[f'\left(\frac{k}{2W}\right)\right]^2,$$

assuming $2TW$ is large.

Hence, approximately,

$$\langle N_o^2\rangle_{AV} = \frac{N}{2TWP'}. \tag{13}$$

Conclusions

When $s(\tau) + N_0$ is small, we may interpret this quantity as $\tau + \Delta\tau$ where $\Delta\tau$ is the error in the measurement of τ due to the noise. Then in the case of white Gaussian noise

$$(\langle\Delta\tau^2\rangle_{AV})^{1/2} = \left(\frac{N}{2WTP'}\right)^{1/2}. \tag{14}$$

If the perfect integrator is replaced by a low-pass filter, then T may, as an approximation, be identified with twice the time constant of the filter. In that case

$$(\langle(\Delta\tau)^2\rangle_{AV})^{1/2} = \left[\frac{(\text{noise power per unit signal bandwidth}) \times (\text{bandwidth of the low-pass filter}).}{\text{average-signal derivative power}}\right]^{\frac{1}{2}}$$

[8] E. T. Whittaker, "On the functions which are represented by the expansions of the interpolation theory," *Univ. of Edinburgh Math. Dept. Res.*, Paper No. 8; 1915.

[9] J. M. Whittaker, "Interpolatory function theory," *Cambridge Tracts in Math. and Math. Phys.*, Tract No. 33, Cambridge University Press, Cambridge, England, 1935.

[10] D. Gabor, "Summary of Communication Theory" from "Communication Theory: Papers Read at Symposium on Applications of Communication Theory, London, September 22–26, 1952" (Willis Jackson, ed.), Academic Press, New York, N. Y., p. 5, 1953.

[11] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Jour.*, vol. 27, p. 379; July, 1948; p. 623, October, 1948.

It will be noted that N/W, or noise per unit bandwidth, is constant with W for white noise whereas the average-signal derivative power increases with W. Hence, a wide signal bandwidth is beneficial provided $s(\tau) + N_0$ is small.

By the radar relation $r = \frac{1}{2}c\tau$, where r is the range of the target, one may use the proposed correlation detector for small signal and noise to determine the range and direction of a target. In such a detector it follows from the above analysis that in the presence of white Gaussian noise a narrow detector bandwidth and a signal with a wide bandwidth would be beneficial. Although filtering alters the spectrum of the noise, it can be shown that a linear filter would not be beneficial. The proof of this fact lies beyond the scope of this paper.

These results have been restricted to the class of real functions $f(t)$, which are stationary time series having finite amplitudes, and whose derivatives exist and are also finite. It is also assumed that $f(t)$ is examined by the correlation detector for a period T of long duration. The Taylor series expansion of $f(t + \tau)$ is assumed to converge.

Optimum Sequential Detection of Signals in Noise*

J. J. BUSSGANG† AND D. MIDDLETON‡

Summary—**A device which performs a sequential test on a mixture of signal and noise is called a *Sequential Detector*. With such a device, two thresholds are introduced, each of which is associated with a terminal decision. The length of the detection process (integration time) is not fixed in advance of the experiment but is a random variable, depending on the progress of the test. An optimum form of such a test exists and is characterized by the fact that detection is performed *on the average* faster than with conventional; *i.e.*, fixed sample size (optimum or non-optimum), devices. The sequential analysis developed by A. Wald is fully applied in this paper, but an important new feature is the treatment of correlated samples and its application to continuous sampling processes.**

In the introduction, the problem is presented within the framework of Wald's Statistical Decision Theory, and the optimum properties of sequential detectors are discussed accordingly. It is pointed out that a sequential detector is defined in terms of *conditional* probabilities and hence its operation is essentially independent of *a priori* information, although the average risk or cost of detection necessarily depends on the *a priori* signal data. The general theory is illustrated with some cases of special interest.

The simplest example of detection involves independent, discrete observations; *e.g.*, the case of a pulsed carrier in normal noise. Here the optimum detector still has the well-known *log I_0* structure, but it is shown that the square law approximation for weak signals requires a bias correction due to the fourth order term. Coherent sequential detection of causal signals in normal noise provides another illustration of the theory. An interesting result is that the probabilities of error do not depend on the shape of the filter, provided the proper computer is used. The use of RC-filtered noise illustrates the treatment of continuous detection processes. Finally, the reduction in minimum detectable signal level resulting from the use of a sequential detector is computed. A third example is the sequential detection of random signals in normal noise. It is shown that, although the optimum computer involves the knowledge of the inverted correlation matrix, the average length of the test does not. Hence a curious result is obtained that in this instance detection can be performed in an arbitrarily short time. The paper concludes with a discussion of the practical necessity of truncating the detection process and exact expressions for the error probabilities of such truncated tests are derived and compared with Wald's original approximations.

* This work was supported in part at Crufts Lab., Harvard University, under a contract with O.N.R., the Signal Corps and the U. S. Air Force. A part of this paper was presented at 1955 Wescon.
† RCA Aviation Systems Lab., Waltham, Mass.
‡ 49 Lexington Avenue, Cambridge, Mass.

I. An Outline of the General Theory of the Binary Sequential Detection Process

Introduction

It has been recognized in recent years that the problem of detection of signals in noise, in its fullest sense, should be viewed as a test of statistical hypotheses [1–8]. If it is important to speed up the detection process, one is led to consider sequential tests, because of their optimum nature. The procedure for such tests is to introduce two thresholds at the detector output such that the signal is declared present if one, and absent if the other of them is exceeded. The length of the detection process (integration time) is not fixed in advance, but is a random variable depending on the progress of the test. Application of sequential tests to detection problems is here termed Sequential Detection. The device which performs the test is called a Sequential Detector. The major feature of these devices is that they minimize the *average* detection time.

In the last ten years, many contributions have been made to the statistical theory of sequential tests [9–14] and the associated analysis has been extensively developed. More recently, sequential procedures have been applied to communication problems [7, 15–22].

To place sequential detection in proper perspective, we begin by locating it within the general domain of Statistical Decision Theory, as applied to reception, following Middleton and Van Meter [23, 24]. Our attention is then directed to simple, binary (*i.e.*, two-decision) detection and an outline of its theory. The mathematical foundations on which it is based are due principally to A. Wald. An application to detection is presented in Part I, which describes the features of the theory of interest to

us here, while Part II illustrates these rather general remarks with important examples encountered in practice. One new feature of the treatment is the handling of correlated samples and continuous sampling processes.

Statistical Decision Problem.

The Statistical Decision Problem has been formulated only recently [25–28] in a form broad enough to include: 1) the state of *a priori* knowledge, 2) losses associated with a false outcome, and 3) costs of experimentation. The problem is considered in relation to a sequence of random variables, say, x_1, x_2, $\cdots$, denoted by $\{x_i\}$, which may represent *e.g.*, successive voltage measurements at the detector output. In general, the joint probability distribution of $\{x_i\}$ is not known exactly. Typically, the form of the distribution function will be known except for a set of parameters $\{\theta\}$. These parameters may be, *e.g.*, the dc level, or the rms value of the voltage examined. Thus while the different x_i may be known to be independent of each other and the distribution gaussian, the associated mean and variance may be unknown. The problem of a statistical decision arises when we are to select one of the mutually exclusive actions, A_1, A_2, $\cdots$, A_k, the degree of preference for which depends on the true value of the set of parameters $\{\theta\}$. These actions may be the initiation of tracking, resumption of search, etc.

As a first step, one may begin by considering a *parameter space* Π which has the dimensionality of the number of unknown parameters (in the example above, two: mean and variance). To each possible set of values of the unknown parameters $\{\theta\}$ there corresponds a point in the space Π. *Before* the experiment begins, the parameter space is divided into zones: π_1, π_2, $\cdots$, π_k such that if the point defined by $\{\theta\}$ were in π_i, the action A_i would be preferred. These zones cannot be overlapping, because the possible actions are mutually exclusive.

The second step, after this subdivision of the parameter space Π, is to determine from *the experiment* in which zone of the parameter space Π the value $\{\theta\}$ falls. Let H_i denote the hypothesis that $\{\theta\}$ lies in π_i. In order to test which hypothesis is correct, one consideres the *sample space* $\sum$ defined by the totality of possible samples. The sample space has the dimensionality of the number of observations in the sample. To each set of possible values of $\{x_i\}$ there corresponds a point in the sample space $\sum$. This space is in turn subdivided into zones σ_1, σ_2, $\cdots$, σ_k such that H_i is to be accepted if, and only if, $\{x\}$ fall in σ_i.

A statistical decision problem can now be defined as the problem of selecting the zones σ_1, $\cdots$, σ_k in the sample space $\sum$, given the zones π_1, $\cdots$, π_k in the parameter space Π. The criterion used in this selection is, to a certain extent, a matter of arbitrary choice for the experimenter. The method advanced by Wald concerns itself with minimizing a certain function, said to be a measure of the *risks* or *costs* involved in making a decision. We now outline the details of this approach.

Minimum Average Risk Criterion.

Suppose, for the sake of simplicity, that the distribution of $\{x\}$ has only one unknown parameter θ; *e.g.*, the signal-to-noise ratio. The parameter space is now a line subdivided into segments. Let $L_i(\theta)$ be the probability of rejecting hypothesis H_i when θ is true, and suppose $C_i(\theta)$ is the loss caused by failure to take action A_i when it should have been taken. If we agree beforehand neither to reward nor to penalize correct decisions, then $C_i(\theta) = 0$ when θ falls outside π_i. The *expected loss* due to a terminal decision is then $\sum_i L_i(\theta)C_i(\theta)$. Now let $n(\theta)$ be the number of observations, when θ is true, and let the *cost* of each observation be c. The *conditional risk* r is then defined by

$$r(\theta) = \sum L_i(\theta)C_i(\theta) + cE[n(\theta)] \tag{1}$$

in which $E[n(\theta)]$ denotes the expected value of $n(\theta)$.

Let g_i be the *a priori* probability that H_i is true ($\sum g_i = 1$). The *average risk* R is accordingly defined by

$$R = \sum_i \sum_j g_j L_i(\theta_j)C_i(\theta_j) + c \sum_j g_j E[n(\theta_j)] \tag{2}$$

in which for the sake of simplicity we let θ assume only discrete values θ_i, $\cdots$, θ_j, $\cdots$ (this restriction is not essential).

Given the cost of experimentation and the loss function C, it may be possible to devise a test which is superior to all other tests in that it entails the smallest risk (usually the smallest *average* risk). Among such optimum tests are: a *Minimax test*, for which the maximum value of the conditional risk r is the least, and a *Bayes test*, for which the *average risk R*, with respect to *a priori* probabilities, is the least. Under certain very weak conditions, a Minimax solution is equivalent to the Bayes solution relative to the least favorable *a priori* distribution [25]. Applications of a general formulation, outlined above, to communication problems have been studied by Middleton and Van Meter [23, 24]. Here, however, we are interested principally in the problem of sequential detection, which is an important aspect of the general theory, and which explicitly introduces the notion of variable sample size or observation periods.

Minimum Risk Formulation of the Detection Problem

In the simplest form of practical interest, detection involves basically a binary (*i.e.*, two-decision) choice between the hypothesis H_0: signal is absent (only noise is present) and the hypothesis H_1: signal as well as noise is present. The parameter for which the test is carried out will be, as a rule, the input signal-to-noise ratio, which is assumed here to remain unchanged during the entire test. The particular value of the input signal-to-noise ratio which is actually the true one will be denoted by a. Our test, then, has to distinguish between the hypothesis H_0 that $a = 0$ and the alternative hypothesis H_1 that a has some specified value a_1. Acceptance of H_1 is known as *an alarm* (action A_1). By analogy, we shall call here the

acceptance of H_0 a *dismissal* (action A_0). The probability of a false alarm (alarm when $a = 0$) is usually designated by α and the probability of a false dismissal (dismissal when $a = a_1$) by β. The quantities α and β are referred to as probabilities of errors of the first and second kind, respectively, just as in the fixed sample size tests. We also term (α, β) the *strength* of the test. The smaller α and β the stronger is the test.

Let p be the *a priori* probability of signal and q be the *a priori* probability of no signal, so that $p + q = 1$. We can now write risks of (1) and (2) as (conditional risk)

$$r(a) = \alpha C_0(a) + \beta C_1(a) + cE[n(a)] \tag{3}$$

and (the average risk)

$$R = q\{\alpha C_0(0) + cE[n(0)]\} + p\{\beta C_1(a_1) + cE[n(a_1)]\}. \tag{4}$$

Notice the relation between the average risk and Siegert's *betting function* [8] defined by

$$S = 1 - (q\alpha + p\beta), \tag{5}$$

which is sometimes specified as a measure of *probable success*. The quantity $1 - S$ is a special case of the average risk discussed above $[c = 0, C_0(0) = C_1(a_1) = 1]$.

Wald and Wolfowitz have shown [10] that whatever the assigned probabilities of error (α, β), costs (C_0, C_1, c) and *a priori* probabilities (p, q), *no test will produce smaller average sample numbers,* $E[n(0)]$ *and* $E[n(a_1)]$, *than the sequential probability ratio test.* Consequently, it follows from (3) and (4) that the sequential probability ratio test minimizes the risks.

The procedure for the sequential probability ratio test is outlined next.

Sequential Test Procedure

A sequential test proceeds in successive stages. At each stage of the experimentation, the sample space $\sum^{(m)}$ is divided into three zones: $\sigma_0^{(m)}$, $\sigma_1^{(m)}$, and $\sigma^{(m)}$; the superscript $^{(m)}$ is used to indicate the m-th stage. The test is terminated in a dismissal if the sample falls in $\sigma_0^{(m)}$, and in an alarm, if it falls in $\sigma_1^{(m)}$. If the sample falls in $\sigma^{(m)}$ the test continues. Here $\sigma^{(m)}$ is called the *zone of indifference* or the *test zone*. It is characteristic of sequential tests that zones of indifference separate *zones of acceptance.* In the basic case considered here, each stage consists of one observation. A terminal decision is reached at some stage n; n is, of course, a random variable, when considered over the ensemble of possible runs. Now let $w_m(\mathbf{x}; a)d\mathbf{x}$ be the probability of obtaining a sample $(x_1, \cdots, x_m)$ when a is true, and consider the *likelihood ratio.*

$$\Lambda_m = pw_m(\mathbf{x}; a_1)/[qw_m(\mathbf{x}; 0)]^{\,1} \tag{6}$$

[1] In the more general case, the signal can assume a continuum of values and the *a priori* distribution of signal, say, $p(a)$ must be given The test procedure is then defined in terms of the ratio

$$\Lambda_m = \left[\int_{\pi_1} w_m(\mathbf{x}; a)p(a)\, da\right] \Big/ [qw_m(\mathbf{x}; 0)] \tag{6a}$$

rather than in terms of (6). For a systematic treatment, including the *a priori* distribution, see [24] (for fixed sample-size).

The division of the sample space is accomplished by selecting two numbers A' and B' $(A' > B')$ and establishing following rules of procedure: The test continues while

$$B' < \Lambda_m < A' \qquad m = 1, 2, \cdots, n - 1, \tag{7}$$

and terminates in acceptance of H_0, if at the nth trial

$$pw_n(\mathbf{x}; a_1) \leq B'qw_n(\mathbf{x}; 0), \tag{8}$$

or in acceptance of H_1, if at the nth trial

$$pw_n(\mathbf{x}; a_1) \geq A'qw_n(\mathbf{x}; 0). \tag{9}$$

In order to establish the connection between A', B' and α, β we consider the probability measure of each side of (8) over the totality of samples leading to acceptance of H_0. This leads to the inequality [9]

$$p\beta \leq B'q(1 - \alpha). \tag{10}$$

Similarly from (9), we obtain

$$p(1 - \beta) \geq A'q\alpha. \tag{11}$$

Let us set $B' = (p/q)B$ and $A' = (p/q)A$. Next, consider the *probability ratio* $\lambda_m = w_m(x; a_1)/w_m(x; O)$ (note that $\Lambda_m = (p/q)\lambda_m$). The test procedure can now be defined on λ_m, as follows: continue sampling while $B < \lambda_m < A$, $m = 1, \cdots, n - 1$; accept H_0 if $\lambda_n \leq B$; and accept H_1 if $\lambda_n \geq A$. A and B are termed the upper and lower boundaries of the test. It is usually assumed that λ_n does not exceed boundaries by an appreciable amount, especially when n tends to be large. Neglecting the so-called "excess over boundaries" [9, p. 44], we have from (10) and (11):

$$A \doteq (1 - \beta)/\alpha \quad \text{and} \quad B \doteq \beta/(1 - \alpha) \tag{12}$$

(in all that follows α and β are assumed less than $\frac{1}{2}$; this assures us that $A > 1$ and $B < 1$).

A simple reasoning in support of (12) is given; *e.g.*, by Mood [29]. If sampling is continuous, the relation (12) is exact. From (12) it follows that

$$\alpha \doteq (1 - B)/(A - B) \text{ and } \beta \doteq B(A - 1)/(A - B). \tag{13}$$

It is important to notice that *the ease with which A and B are related to* α *and* β *is entirely due to the fact that the detector used is a probability ratio detector.* It is also important to notice that the sequential test is set in terms of conditional probabilities (probability of error *if* signal is present and probability of error *if* signal is absent). Thus the sequential test can be constructed *without using* p and q; it is *independent* of *a priori* probabilities. The average risk R, of course, always depends on *a priori* probabilities. Hence p and q are needed when a_1 is defined as the least value of signal-to-noise ratio for which the average risk (given p, q, and the costs) does not exceed some specified number. This number would be the maximum acceptable average risk and a_1 would be correspondingly called in this connection the *minimum signal-to-noise ratio.* For any signal-to-noise ratio larger than a_1, the resulting average risk would be acceptable.

The question of whether a sequential test will terminate is of fundamental importance. Stein [12] showed that all moments of n will be finite if observations are independent. Moreover, Wald has pointed out [9, p. 43] that the *probability is unity that the test will eventually terminate, for a very large class of joint distributions when observations are not independent.* That this is so, granted Stein's result, follows from physical considerations. For any band-limited process (with no dc or purely periodic component) it is possible by taking observations far enough apart in time to produce independent samples. In addition, it may be pointed out that if the expected length of a sequential test of strength (α, β) were not finite, a classical test of this strength would not be realizable. This is a direct consequence of the optimum nature of sequential tests. It must be recognized, however, that some individual sequential tests may take a very long time to terminate. If such a feature were undesirable, one would have to resort to some new rule for terminating after a certain stage has been reached. Among such possible solutions is the truncation procedure discussed in Part II.

Handling of A-Priori Information

The setting up of a problem depends on the amount of *a priori* information available. Three general situations are recognized:

Case 1: Value of Signal Known Exactly. Suppose the signal is known to occur only at a specific value, if at all. Then we are dealing with two simple alternatives, for which the sequential test procedure is that described above. This is the simplest possible situation.

Case 2: Distribution of Signals Known Exactly. Suppose $p(a)$, the *a priori* distribution of signals, is known exactly. Case 1 is a specialized form of Case 2 with $p(a) = p\delta(a - a_1)$. The test procedure is now defined in terms of the weighted probability of a sample; *i.e.*, $\int_{\pi_1} w_n(\mathbf{x}, a)\, p(a)\, da\, d\mathbf{x}$, where π_1 includes all the possible non-zero signal values; (see footnote reference 1). In place of the probability of false dismissals $p\beta$, we now have an average probability of false dismissals $\bar{\beta}$, defined by

$$\bar{\beta} = \int_{\pi_1} da\, p(a) \int_0^{\infty} p(n \mid d_0; a) \int_{\sigma_0(n)} w_n(\mathbf{x}; a)\, d\mathbf{x} \tag{14}$$

in which

$$q + \int_{\pi_1} p(a)\, da = 1. \tag{15}$$

and $p(n|do; a)$ is the conditional distribution of n if H_0 is accepted and a is true.

From considerations similar to those in the preceding section, it can be shown that A' and B' are approximately given by $A' = (1 - \bar{\beta})/(q\alpha)$ and $B' = \bar{\beta}/[q(1 - \alpha)]$. The boundaries of the test are still determined by α and $\bar{\beta}$ in a simple fashion. Whether the distribution of the sample size can be derived will depend entirely on how complex is the form of Λ_m for the specific $p(a)$.

Case 3: Distribution of Signals Known Incompletely. This is perhaps the most common case in a practical situation. The procedure here is to select an α and to fix β at a suitable signal level ($a = a_1$). Then one constructs the test as if one were dealing with a simple alternative case. With this method only the choice of a_1 involves any *a priori* knowledge. Notice that although the test is now constructed as if signal-to-noise ratios 0 and a_1 were the only possible alternatives, the actual signal-to-noise ratio, say a, will be in general different from either of these two alternatives.

The probability of a terminal decision "to dismiss" is a function of both a and a_1, and is denoted by $L(a)$.

Before we examine $L(a)$, let us emphasize that sequential tests pose similar problems with respect to the handling of and the requirements for *a priori* information as do classical tests; *e.g.*, both Minimax and Bayes sequential tests are possible under suitable circumstances.

Operating Characteristic Function (OCF)

The probability $L(a)$ of accepting H_0 at the end of a test is termed by Wald the Operating Characteristic Function (OCF) of the test [9]. As a rule, $L(a)$ decreases as a increases. Because we consider a test which is certain to terminate, the probability of accepting H_1 is given by $1 - L(a)$. For independent observations, Wald has developed a method of finding $L(a)$ with the aid of a parametric equation. This method can be extended to the case of correlated samples.

Suppose we are testing hypothesis H_1 that $w_m(\mathbf{x}; a_1)$ is the distribution of $\mathbf{x}$ (signal present) against hypothesis H_0 that the distribution is $w_m(\mathbf{x}; O)$ (signal absent). Let the upper threshold be A and the lower B. The (conditional) probability ratio of this test is

$$\lambda_m = w_m(\mathbf{x}; a_1)/w_m(\mathbf{x}; 0). \tag{16}$$

Following procedure described in last paragraph, we accept H_0 (declare signal absent) whenever $\lambda_m \leq B$. Construct now λ_m^h, where $h = h(a, a_1)$ satisfies the condition

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \lambda_m^h w_m(\mathbf{x}; a)\, d\mathbf{x} = 1; \tag{17}$$

h is then a parameter depending on the distribution of the observed variables. From the last equation it follows that $f_m(\mathbf{x}) = \lambda_m^h w_m(\mathbf{x}; a)$ is, quite formally, some distribution of $\mathbf{x}$. If we next consider a new sequential test, whose object is to decide whether $f_m(\mathbf{x})$ or $w_m(\mathbf{x}; a)$ is the true distribution of $\mathbf{x}$, the probability ratio of this new test is λ_m^h. Letting the boundaries of the new test be C and D, we see that the probability of declaring $w_m(\mathbf{x}; a)$ the true distribution when it indeed is true, is from (13)

$$1 - \alpha' = (C - 1)/(C - D). \tag{18}$$

Let us now select $C = A^h$ and $D = B^h$ $(h > 0)$; then the new test has the property that $w_m(\mathbf{x}; a)$ is declared true $(\lambda_m^h \leq B^h)$ whenever H_0 is declared true in the original test $(\lambda_m \leq B)$. Hence the probability of accepting H_0 when $w_m(\mathbf{x}; a)$ is true is, from (18), $(A^h - 1)/(A^h - B^h)$. The

same relation can be shown to hold for h negative. Hence we get the Operating Characteristic Function

$$L(a) = (A^h - 1)/(A^h - B^h), \tag{19}$$

in which h satisfies the relation

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [w_m(\mathbf{x})a_1)/w_m(\mathbf{x};\,0)]^{h(a,a_1,m)} w_m(\mathbf{x};\,a)\,d\mathbf{x} = 1. \tag{20}$$

For independent observations, we have $w_m(\mathbf{x};\,a) = [w_1\,(x;\,a)]^m$ and (20) reduces to the form

$$\int_{-\infty}^{\infty} [w_1(x;\,a_1)/w_1(x;\,0)]^{h(a,a_1)} w_1(x;\,0)\,dx = 1. \tag{21}$$

It is easy to verify that $h(O) = 1$ and $h(a_1) = -1$. Also we have $L(O) = 1 - \alpha$ and $L(a_1) = \beta$, as required.

The value of a at which $h = 0$ will be denoted by a'. A limiting process has to be applied at this point:

$$L(a') = \log A/\log(A/B) \tag{22}$$

$$\therefore \left(\frac{dL}{dh}\right)_{a=a'} = (\log A)(\log B)/[2\log(A/B)]. \tag{23}$$

When $\alpha = \beta$ (and hence $A = 1/B$), we get from (19)

$$L(a) = 1/(1 + A^{-h}). \tag{24}$$

The OCF of the test is not only important in its own right but it is also needed for the computation of risks and most important of all, it allows the evaluation of the Average Sample Number of the test.

Average Sample Number (ASN)

The sequential test procedure has been discussed in terms of the probability ratio $\lambda_m = w_m(\mathbf{x};\,a_1)/w_m(\mathbf{x};\,0)$. It is convenient now to consider $\log \lambda_m$, which we denote by Z_m; *i.e.*,

$$Z_m = \log[w_m(\mathbf{x};\,a_1)/w_m(\mathbf{x};\,0)]. \tag{25}$$

The test procedure can be stated in terms of Z_m, as follows: Construct Z_m at each state of the experiment. Continue testing while

$$\log B < Z_m < \log A, \qquad m = 1, 2, \cdots, n-1. \tag{26}$$

Terminate by accepting H_0 when $Z_n \leq \log B$. Terminate by accepting H_1 when $Z_n \geq \log A$.

Since the average value of Z_n, say $\overline{Z}_n$, must be equal to the value of the bounds, weighted by the probability of reaching them, we have

$$\overline{Z}_n \doteq L(a)\log B + [1 - L(a)]\log A. \tag{27}$$

Now for independent observations $Z_m = \sum_{i=1}^m z_i$, where

$$z_i = \log[w_1(x_i;\,a_1)/w_1(x_i;\,0)]. \tag{28}$$

Let $\bar{n}$ denote the Average Sample Number (average number of observations required to terminate the test); then for independent observations $\overline{Z}_n = \bar{n}\bar{z}$, and by virtue of (27), one gets Wald's result [9] for the average sample size when the input signal-to-noise (rms amplitude) ratio is a

$$\overline{n(a)} = \{L(a)\log B + [1 - L(a)]\log A\}/\overline{z(a)}; \tag{29}$$

($\overline{n(a)}$ is used as an alternative notation for $E[n(a)]$.) For correlated observations it may be still possible to obtain a simple relation for $\bar{n}$ from (27). The condition is that $\overline{Z}_n$ be expressible as a linear function of $\bar{n}$. Specific examples are taken up in Part II. If we let

$$g(\alpha, \beta) \equiv (1 - \alpha)\log B + \beta\log A, \tag{30}$$

we can express $\overline{n(0)}$, the ASN in absence of signal, and $\overline{n(a_1)}$, the ASN in presence of signal, as

$$\overline{n(0)} = g(\alpha, \beta)/\overline{z(0)}, \quad \text{and} \quad \overline{n(a_1)} = -g(\beta, \alpha)/\overline{z(a_1)}. \tag{31}$$

This means that $\overline{n(a_k)}$ $(k = 0, 1)$ can be split up into two factors, one of which depends only on the probabilities of error (α, β) and the other only on the distribution function of the observed variable and on the value a_1. Hence, if two detectors of the same strength (α, β) are to be compared by their required sample sizes, only the quantities $\bar{z}$ need be examined. The auxiliary function $g(\alpha, \beta)$ is plotted on Fig. 1.

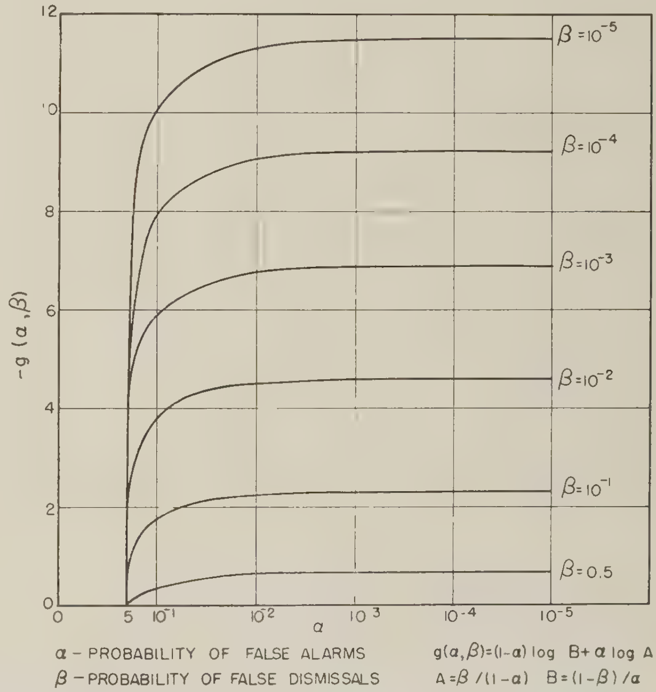


Fig. 1—The auxiliary function $g(\alpha, \beta)$.

Notice that at $a'(h = 0)$, the numerator of (29) becomes zero, since $\overline{z(a')} = 0$.[2] The relation (29) is therefore indeterminate at $a = a'$. In order to find $\overline{n(a')}$, the ASN when $a = a'$, we consider the expected value of Z_n^2. At termination, Z_n^2 must be equal to the square of either one or the other bound. Taking account of probabilities with which the test terminates in either decision, we get

$$\overline{Z_n^2} = L(\log B)^2 + (1 - L)(\log A)^2. \tag{32}$$

[2] Eq. (21) defining h can also be written $\int_{-\infty}^{\infty} e^{zh}w(z)\,dz = 1$. Differentiating both sides with respect to h and setting $h = 0$, $(a = a')$, we get $\overline{z(a')} = 0$.

Now for independent observations, if $\bar{z} = 0$, we have

$$\overline{Z_n^2} = \bar{n}\overline{z^2}, \tag{33}$$

so that recalling (22) we obtain

$$\overline{n(a')} = -(\log A)(\log B)/\overline{z^2(a')}, \tag{34}$$

in which $\log B$ is negative ($B < 1$), and consequently $\bar{n} > 0$.

In the important special case of equal probabilities of error ($\alpha = \beta$; hence $A = 1/B$), relation (29) becomes

$$\overline{n(a)} = -(\log A)[\tanh(\tfrac{1}{2}h \log A)]/\overline{z(a)}. \tag{35}$$

[The negative sign is accounted for by the fact that h and $\bar{z}$ always have opposite signs.]

For small α and β one gets simply the limiting forms:

$$\overline{n(a_1)} \to (\log A)/\overline{z(a_1)} \quad \text{and} \quad \overline{n(0)} \to (\log B)/\overline{z(0)}. \tag{36}$$

Continuous Detection

In many communication problems detection is to be carried out as a continuous process; *i.e.*, instead of a series of discrete sample values the data is observed continuously. The theory of sequential detection can be extended in a natural manner to apply to such continuous sampling, if the noise is Gaussian.

Consider observations spaced uniformly at times Δt apart. The expression for Z_m will in general involve a summation, which because of correlation between observations depends on Δt. One can let $\Delta t \to 0$, keeping at the same time $m\Delta t = T$ finite, to obtain a detector output Z_T. The rules of procedure stated for Z_m in the last section now apply to Z_T. Usually Z_T requires an integration (in place of summation), in which T appears as the upper bound of integration and 0 as the lower one. Suppose that the test terminates at some time $\mathfrak{J}$, corresponding to $n\Delta t$. Then $\mathfrak{J}$ is a random variable, inasmuch as n is. The probability $L(a)$ of accepting H_0 (no signal) at the end of the test can still be introduced. One obtains the fundamental relation

$$\overline{Z_{\mathfrak{J}}} = L \log B + (1 - L) \log A. \tag{37}$$

From this relation the Average Sample Length $\bar{\mathfrak{J}}$(cf. with $\bar{n}$) can be obtained, provided $\overline{Z_{\mathfrak{J}}}$ is expressible in terms of $\bar{\mathfrak{J}}$.

An alternative method of treating correlated samples of Gaussian noise involves the choice of a proper linear transformation leading to a new set of variables $y_1, y_2, \cdots, y_m$, such that the new variables are independent and all follow the same distribution. Since the test is sequential, y_i are made independent of all x_k for $k > i$; this amounts to having a triangular transformation matrix and is closely related to the problem of inverting the original matrix of correlations. An example of this technique is given in the second section of Part II.

Under certain conditions, the entire distribution of $\mathfrak{J}$, rather than just $\bar{\mathfrak{J}}$, can be obtained [21, 22]. Of particular interest is, of course, $\overline{\mathfrak{J}^2}$, the second moment of $\mathfrak{J}$, which provides a measure of dispersion. $\mathfrak{J}$ is in effect the first passage time of a random walk with absorbing barriers and in addition to the Sequential Analysis can be studied by going back to the original differential equations of the process (*e.g.* [30, 31].

II. Examples of the Optimum Sequential Detection of Signals in Noise

We shall consider here specific distributions of the observed variable corresponding to some typical detection problems. The examples selected are: incoherent detection of a sinewave in normal noise; coherent detection of a causal signal in normal noise; and detection of a normal noise "signal" in normal random noise.

Incoherent Detection of a Sinewave in Noise:

Let the random variable under inspection be the envelope of a narrow-band noise and an additive sinewave at the center frequency of the noise band. The noise is assumed to have a Gaussian distribution function with a zero mean. Under these conditions the distribution of signal-plus-noise envelope is given by the well-known expression [32]

$$\begin{aligned} W(X; A) &= \frac{X}{\psi} e^{-(X^2+A^2)/(2\psi)} I_0(XA/\psi), &\quad X \geq 0 \\ &= 0, &\quad X < 0, \end{aligned} \tag{38}$$

where X is the amplitude of the output envelope, ψ is the mean square value of noise, A is the peak amplitude of the sinewave and $I_0(u)$ is the modified Bessel function of the first kind, zeroth order. It is convenient to change the notation by letting

$$x = X/\sqrt{2\psi} \quad \text{and} \quad a = A/\sqrt{2\psi}; \tag{39}$$

a is then simply the square root of the signal-to-noise power ratio. The probability density of x is accordingly

$$\begin{aligned} w(x; a) &= 2xe^{-a^2-x^2} I_0(2ax), &\quad x \geq 0 \\ &= 0, &\quad x < 0. \end{aligned} \tag{40}$$

Note that to get (38) an average over phase has been performed corresponding to the *a priori* knowledge that the distribution of phase angle is uniform (see, *e.g.*, [33, 34]). In what follows the observed variable is x, and the observations are assumed independent of one another. The sequence of n observations (voltage measurements) is referred to as a sample of length n. The distribution of x has an unknown parameter a: if $a = 0$, the signal is absent; if $a > 0$, a signal is present. We define a random variable z as the logarithm of the probability ratio, $z = \log [w(x; a_1)/w(x; 0)]$, so that from (40)

$$z = -a_1^2 + \log I_0(2a_1x). \tag{41}$$

The probability ratio detector constructs z_i, corresponding to each sample value x_i of the received wave. A classical, fixed sample test requires the construction of the same function z [7]. The difference between classical and sequential test procedures arises when instead of adding up a fixed number of z_i and checking the sum against *a single threshold*, we now form $\sum_{i=1}^{m} z_i$, observation by

observation, checking this sum against *two thresholds* at each stage m (see Part I, Sequential Test Procedure).

For weak signals a power series expansion can be made in (41) and we obtain

$$z \doteq -a_1^2 + a_1^2x^2 - \tfrac{1}{4}a_1^4x^4 + 0(a_1^6x^6). \qquad (42)$$

In the past, it has been customary to retain only the constant and the term in x^2.

It has been pointed out recently [20–22] that the contribution of the fourth order term is significant in computing $\bar{z}$, the first moment of z. However, it is still possible to approximate the optimum detector by a square law operation on x, provided the bias be modified to include $\frac{1}{2}a_1{}^4\overline{x^4}$, the expected value of the fourth order term. The distribution function (40) leads to $\overline{x^2} = 1 + a^2$ and $\overline{x^4} = 2!\,(1 + 2a^2 + \frac{1}{2}a^4)$ (see *e.g.*, [8, 32]). Hence the weak signal approximation to the detector of (41) is *not*

$$z \doteq -a_1^2 + a_1^2x^2 \qquad (43)$$

but it is

$$z \doteq -a_1^2(1 + \tfrac{1}{2}a_1^2) + a_1^2x^2. \qquad (44)$$

Thus the correct expression for $\bar{z}$ is

$$\overline{z(a)} = a^2a_1^2 - \tfrac{1}{2}a_1^4, \qquad a,\ a_1 \ll 1, \qquad (45)$$

leading to $\overline{z(0)} = -\frac{1}{2}a^4$, and $\overline{z(a_1)} = \frac{1}{2}a_1^4$ and not $\overline{z(0)} = 0$ and $\overline{z(a_1)} = a_1^4$ as has sometimes been assumed in the past. Notice that if $\overline{z(0)} = 0$ were true, it would lead to an infinite Average Sample Number in absence of signal, since we have seen that $\bar{n}$ is inversely proportional to $\bar{z}$. An infinite ASN would, of course, contradict the theorem that sequential tests terminate in a finite number of observations with probability 1 (see the Sequential Test Procedure section).

We emphasize the importance of the correction bias due to the fourth order term because this point does not seem to have been generally appreciated in the treatment of nonsequential detectors. The exact value of bias is important if the bias is used as an independent variable. In many cases bias is eliminated between the expressions for errors of the two kinds. In such cases, the use of the square law term alone is good enough, and it does not matter whether or not the fourth order term is used. The overall conclusion is that depending on the statistics of the variable under test, terms higher than the square may be important and should in each case be carefully examined.

A detector constructed according to (41) is optimum only for a specific signal-to-noise ratio a_1. This is equally true of sequential and of nonsequential operations. The effect of the true signal-to-noise ratio on the performance can be determined from the Operating Characteristic Function (OCF) which gives the probability of H_0 being accepted at the end of a test (see Section on OCF in Part I). The OCF is, we recall, given by $L(a) = (A^h - 1)/(A^h - B^h)$, where h is now defined by

$$2\int_0^\infty xI_0(2ax)[I_0(2a_1x)]^h e^{-x^2}\,dx = e^{a^2+ha_1^2}. \qquad (46)$$

While the general solution of (46) presents problems in the weak signal case appropriate to threshold reception, we obtain approximately $h = 1 - 2\,(a/a_1)^2$. Some typical OCF curves are shown in Fig. 2. The point at which the slope is maximum may properly be called the point of greatest ambiguity as to the outcome of the test.

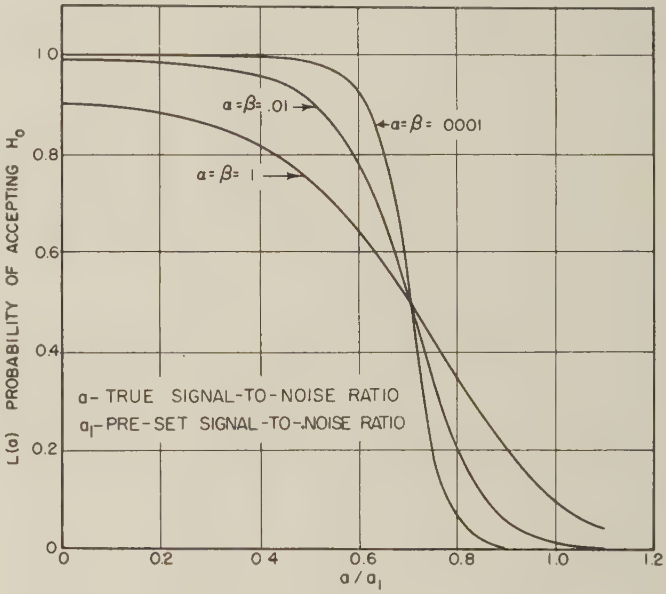


Fig. 2—OCF of the incoherent detector; weak signals. (a, $a_1 \ll 1$).

Once the OCF is known, the ASN can be computed from (29). The additional quantity needed is $\overline{z(a)}$, and so we have

$$\overline{z(a)} = -a_1^2 + \int_0^\infty [\log I_0(2a_1x)]2xe^{-x^2-a^2}I_0(2ax)\,dx. \qquad (47)$$

When $a = 0$, this last expression reduces to the form

$$\overline{z(0)} = -2a_1^2\int_0^\infty [I_2(2a_1x)/I_0(2a_1x)]x\exp(-x^2)\,dx. \qquad (48)$$

For weak signals ($a_1 \ll 1$), we obtain [22]

$$\overline{z(0)} = -\frac{1}{2}a_1^4\left(1 - \frac{4}{3}a_1^2 + \frac{11}{4}a_1^4 - \cdots\right). \qquad (49)$$

For strong signals, we get [22] on the other hand

$$\overline{z(a_1)} = -a_1^2\left(0.45 - \frac{1}{a_1} + \frac{0.49}{a_1^2} - \cdots + 2\ln a_1\right) - 0.178. \qquad (50)$$

From (47) one also finds that for weak signals ($a_1 \ll 1$)

$$\overline{z(a_1)} = \frac{1}{2}a_1^4\left(1 - \frac{2}{3}a_1^2 + \frac{3}{4}a_1^4 - \frac{7}{5}a_1^6 + \cdots\right). \qquad (51)$$

In Fig. 3, $\overline{z(a_1)}$ and $|\overline{z(0)}|$ are plotted. It is rather remarkable to notice that $|\overline{z(0)}|$ and $\overline{z(a_1)}$ are approximately equal over the entire range of a_1. This relation is not obvious from a simple inspection of (47) and on the basis of (30) can be taken to mean that when there is no signal ($a = 0$), a test of strength (α, β) will last, on the average, approximately as long as a test of strength (β, α)

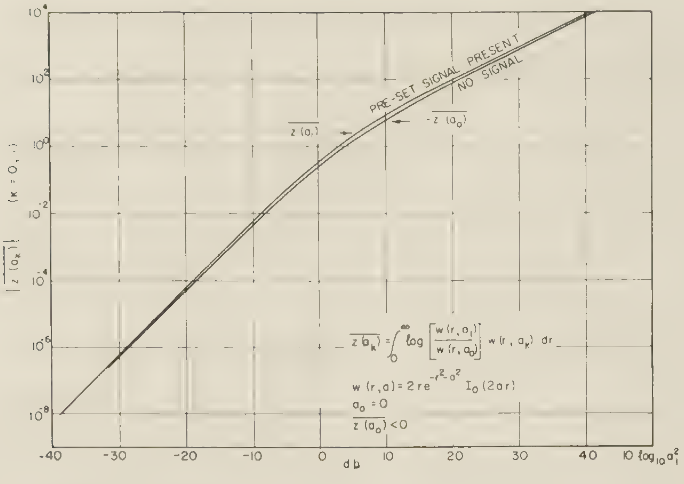


Fig. 3—Expected value of z for the incoherent detector.

when signal appears at the preset level ($a = a_1$). The relation $-\overline{z(0)} \doteq \overline{z(a_1)}$ seems characteristic of sequential tests in general and holds exactly when the observed variable has Gaussian statistics (see below).

For a more complete understanding of the mechanism of a sequential test it is not enough to know $\bar{n}(0)$ and $\bar{n}(a_1)$, but it is also important to examine the ASN as a function of a. A family of such curves is shown in Fig. 4.

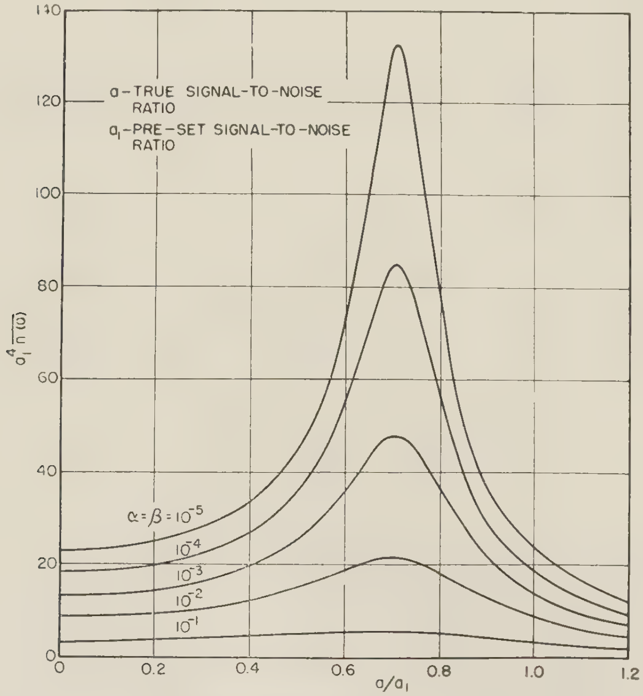


Fig. 4—ASN ($\times a_1^4$) of the incoherent detector vs a/a_1; weak signals. (a, $a_1 \ll 1$).

The most striking feature is the peak of $\bar{n}$ at a value between $a = 0$ and $a = a_1$. This peak can be explained by the fact that there is no pronounced tendency to cross either boundary. The ASN increases as the probability of errors is reduced. The essence of sequential procedure is that we save on the average number of observations at the price of randomizing the sample length n. It is of interest therefore to know the variance of n. We denote the variance of n by $\sigma_n^2(a)$, in which a is the true value of the signal-to-noise ratio. Let the variance of z be denoted by $\sigma_z^2(a)$. It can be shown [22] for small α and β, and consequently large $\bar{n}$, that the variance of n is approximately given by

$$\sigma_n^2(0) \doteq (\log B)\sigma_z^2(0)/\overline{z(0)}^3 \tag{52}$$

when no signal is present, and by

$$\sigma_n^2(a_1) \doteq (\log A)\sigma_z^2(a_1)/\overline{z(a_1)}^3. \tag{53}$$

when signal is present. The larger the variance of n, the more saving on the Average Sample Number. For small variance there is little saving, since there is then little difference between a sequential and a nonsequential test.

Coherent Detection of Signals in Normal Random Noise.

The envelope detection which was discussed above requires knowledge of the amplitude but not of the detailed (RF) phase structure of the signal. For coherent detection, on the other hand, the fine-structure of the signal must be known. Treatment of optimum fixed sample size detectors is available in several publications *e.g.*, [5–7, 34]. We shall now discuss sequential, coherent detectors.

Suppose a causal signal voltage is given as a function of time by $As(t)$. Let the value of s at the time t_i, corresponding to the ith observation, be denoted by s_i; *i.e.*, $s(t_i) = s_i$. We assume $s(t)$ normalized in such a fashion that

$$\lim_{N\to\infty} (1/N) \sum_{i=1}^{N} s_i^2 = 1; \tag{54}$$

hence A^2 measures the mean square value of the observations (as $N \to \infty$) and $s(t)$ gives the normalized waveform. When the signal is absent, $A = 0$, and when signal is present $A = A_1$. Let us denote the amplitude of a mixture of signal and noise by X. The noise is assumed Gaussian with a zero mean and an rms value ψ. In this example, detection corresponds to a statistical test for the mean of X, whose value on the ith observation is either zero (H_0) or A_1 (H_1). In the simplest case, all s_i are equal. More generally, the s_i differ from each other but assume prescribed values. When the s_i are not all equal, a difficulty peculiar to sequential detection is introduced: in averaging over the test length, the upper limit is itself a random variable. Such averaging occurs when the Average Sample Number (ASN) is to be computed. If the signal is periodic, as is the case in some practical situations, the difficulty can be removed, provided tests last long enough to render end-effects negligible. The test itself can always be set up to have a certain strength; this is assured by the choice of thresholds. Analytic difficulties arise only when the ASN is to be computed.

Let us next perform a normalization indicated by $x = X/\psi^{\frac{1}{2}}$ and $a = A/\psi^{\frac{1}{2}}$, so that a is the rms (input) signal-to-noise ratio. We consider x as the observed variable and a as the unknown parameter. The null hypothesis H_0, of no signal, corresponds to $a = a_0 = 0$.

The alternative hypothesis H_1, of signal at threshold level, corresponds to $a = a_1$. The m-dimensional distri-

bution function of a sample $\mathbf{x}(x_1, \cdots, x_m)$ of signal-plus-noise under hypothesis H_k $(k = 0, 1)$ is

$$w_m(\mathbf{x}; a_k\mathbf{s}) = (2\pi)^{-m/2}[\det(\sigma_{ij})]^{-1/2} \cdot \exp\left[-\frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}(x_i - a_k s_i)(x_j - a_k s_j)\right] \quad (55)$$

in which $\mathbf{s}$ is used to indicate the values $s_1, \cdots, s_m$ of signal. The determinant of the matrix of covariances is denoted by $\det(\sigma_{ij})$ and elements of the inverse matrix by σ^{ij}; note that $\det(\sigma_{ij}) = [\det(\sigma^{ij})]^{-1}$. Because of the chosen normalization, we have $\sigma_{ii} = 1$ for all i. The logarithm of the probability ratio $w_m(\mathbf{x}; a_1\mathbf{s})/w_m(\mathbf{x}; 0)$ is

$$Z_m = -\frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}(a_1^2 s_i s_j - a_1 s_j x_i - a_1 s_i x_j). \quad (56)$$

If the signal does not vary from observation to observation $(s_i = s_j = 1$, etc), one has

$$Z_m = -\frac{1}{2}\sum_{i,j}^{m}\sigma^{ij}[a_1^2 - a_1(x_i + x_j)]. \quad (57)$$

An important illustration of the structure of a probability ratio detector is supplied by noise which has an exponential autocorrelation function, $r(t) = \exp(-|t|)$. Such an autocorrelation function is encountered at the output of an RC-filter through which "white" noise is passed. It also corresponds to the envelope of narrow-band RLC noise. If observations are taken at equal intervals, say D, the normalized correlation coefficient between the ith and the jth observation depends only on $|i - j|$, that is $\sigma_{ij} = \rho^{|i-j|}$ where $\rho = \exp(-\gamma D)$. The inverse matrix for this case has been obtained by Reich and Swerling [6] and is relatively simple: $\sigma^{11} = \sigma^{mm} = 1/(1 - \rho^2)$, $\sigma^{ii} = (1 + \rho^2)/(1 - \rho^2)$ $(i \neq 1, m)$, $\sigma^{i,i+1} = \sigma^{i,i-1} = -\rho/(1 - \rho^2)$, all other $\sigma^{ij} = 0$. Substituting these values for σ^{ij} in (56), we get

$$Z_m = -\frac{1}{2}(1 - \rho^2)^{-1}\left\{\sum_{i=2}^{m-1} a_1^2(1 + \rho^2)s_i^2 - a_1^2\rho s_i(s_{i+1} + s_{i-1}) + 2a_1(1 + \rho^2)s_i x_i - 2a_1\rho x_i(s_{i+1} + s_{i-1}) + a_1^2(s_1^2 + s_m^2) - a_1\rho(s_1 s_2 + s_{m-1}s_m) + 2a_1(s_1 x_1 + s_m x_m)\right\}. \quad (58)$$

Introducing the standard difference operators $\Delta s_i = s_{i+1} - s_i$ and $\Delta^2 s_i = s_{i+2} - 2\,s_{i+1} + s_i$ and neglecting the end-effects (terms in $i = 1$ and $i = m$ give insignificant contributions if m is large), we are left with

$$Z_m = -\frac{1}{2}(1 + \rho)^{-1}\left[(1 - \rho)\sum_{i=2}^{m-1} s_i(a_1^2 s_i - 2a_1 x_i) - \frac{\rho}{1 - \rho}\sum_{i=2}^{m-1}\Delta s_i(a_1^2 s_i - 2a_1 x_i)\right]. \quad (59)$$

For a small spacing between observations, $\rho = 1 + \gamma\,\Delta t + \cdots$; hence Z_m can be approximated by

$$Z_m \doteq -\frac{1}{4}(1/\gamma)\sum_{i=2}^{m-1}(a_1^2 s_i - 2a_1 x_i)(\gamma^2 s_i - \Delta^2 s_i)\Delta t. \quad (60)$$

Transition to the continuous case is effected by letting $\Delta t \to 0$ and setting the duration T of the test equal to $m\Delta t$ in the limit. We find that

$$Z_T \doteq -\frac{1}{4}(1/\gamma)\int_0^T [a_1^2 s(t) - 2a_1 x(t)][\gamma^2 s(t) - \ddot{s}(t)]\,dt, \quad (61)$$

exclusive of end-effects, which are ignorable if $T \gg \gamma^{-1}$. In particular for a sinewave of radial frequency ω_0, $\ddot{s} = -\omega_0^2 s$, so that one has

$$Z_T = \frac{1}{4}\gamma(1 + \omega_0^2\gamma^{-2})\int_0^T [a_1^2 s^2(t) - 2a_1 x(t)s(t)]\,dt. \quad (62)$$

The detection procedure in this example is as follows: Supply the detector with the signal waveform $s(t)$. Continue integrating the quantity $[a_1^2 s^2(t) - 2a_1 x(t)s(t)]$, or, in effect, crosscorrelating $x(t)$ and $s(t)$. Detection terminates in an alarm or a dismissal, respectively, if either of the two limits $(4\gamma \log A)/(\gamma^2 + \omega_0^2)$ or $(4\gamma \log B)/(\gamma^2 + \omega_0^2)$ is exceeded.

Another approach to the problem of designing optimum sequential detectors for correlated samples with a Gaussian distribution is to choose a proper linear transformation leading to a new set of variables $y_1, \cdots, y_m$, such that the new variables are independent and are all made to follow the same distribution. Moreover, the y_i are made independent of all the x_k for $k > i$. Using again the example of RC-filtered noise, consider the transformation

$$\begin{aligned} y_1 &= x_1(1 - \rho^2)^{1/2} \\ y_2 &= x_2 - \rho x_1 \\ &\vdots \\ y_m &= x_m - \rho x_{m-1} \end{aligned} \quad (63)$$

in which $\rho = \overline{x_i x_{i+1}} - \overline{x_i}\,\overline{x_{i+1}}$, $\overline{x_i^2} - \overline{x_i}^2 = 1$ and $\overline{x_i} = as_i$. Clearly, we have $\overline{y^2} - \overline{y}^2 = 1 - \rho^2$. The distribution of any y_i (except y_1) is

$$w(y_i; as_i) = [2\pi(1 - \rho^2)]^{-1/2} \cdot \exp\{-[y - a(s_{i+1} - \rho s_i)]^2/2(1 - \rho^2)\}. \quad (64)$$

The logarithm of the probability ratio, $\xi_i = w(y_i; as_i)/w(y_i; 0)$ is then given by

$$\xi_i = \tfrac{1}{2}[2a_1 y(s_{i+1} - \rho s_i) - a_1^2(s_{i+1} - \rho s_i)^2]/(1 - \rho^2). \quad (65)$$

Substituting for y_i in terms of x, we obtain (except for ξ_1)

$$\xi_i = \tfrac{1}{2}(s_{i+1} - \rho s_i)[2a_1(x_{i+1} - \rho x_i) - a_1^2(s_{i+1} - \rho s_i)]/(1 - \rho^2). \quad (66)$$

The detection is performed by checking at each stage whether or not $\sum_i \xi_i$ exceeds one of the two usual bounds, $\log A$ or $\log B$. As expected, $\sum_i \xi_i$ checks with Z_m in (58). The operation of inverting a matrix, required by the first method, and the operation of finding a triangularizing transformation are, of course, intimately related (inversion of a large class of correlation matrices is discussed in [35]). Either one or the other has to be carried out in order to determine the form of an optimum detector.

The analysis of a sequential detector involves finding the OCF and the ASN. The OCF is obtained from the standard relation $L(a) = (A^h - 1)/(A^h - B^h)$, in which h is the nonzero root of the equation

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [w_m(\mathbf{x}; a_1\mathbf{s})/w_m(\mathbf{x}; 0)]^{h(a,a_1,m)} w_m(\mathbf{x}; a\mathbf{s})\, d\mathbf{x} = 1. \tag{67}$$

Substituting from (55), we find that the condition on h is

$$\exp\left\{-\sum_{i,j}^{m} [\sigma^{ij} s_i s_j a_1 h(a_1 - 2a - ha_1)/2\right\} = 1. \tag{68}$$

This means that h satisfies the simple expression

$$h = 1 - 2a/a_1, \tag{69}$$

which is independent of sample size.

This result shows that for normal noise the *Operating Characteristic Function of a sequential test is independent of the correlation matrix* $[\sigma_{ij}]$, *of the specified signal waveform* s_i, *and of the number of the observations* m. Since the correlation matrix is not involved, one concludes that observations can be allowed to be spaced arbitrarily close and hence that, in the limit, the same result carries over to the continuous case.

In Figs. 5, 6, and 7, the OCF's of the coherent detector are plotted. The OCF is important in that it indicates the performance of the detector when the *actual* signal-to-noise ratio (a) is not as large as the *preset* signal-to-noise ratio (a_1) for which the detector is to have strength (α, β). The OCF is also used in determining the Average Sample Number (ASN). The basic expression is here

$$\bar{Z}_n = L(a) \log B + [1 - L(a)] \log A. \tag{70}$$

For the coherent detector, from (56), we get

$$\bar{Z}_n = E\left[-\frac{1}{2}\sum_{i,j}^{n} \sigma^{ij}(a_1^2 s_i s_j - a_1 s_j x_i - a_1 s_i x_j)\right]. \tag{71}$$

Whether an explicit expression for $\bar{n}$ can be found will depend on the specific waveform s_i, and on the structure of the correlation matrix $[\sigma_{ij}]$.

Consider the case in which the signal value is constant at each observation ($s_i = 1$). We then have

$$\bar{Z}_n = -\frac{1}{2}(a_1^2 - 2aa_1)E\left(\sum_{i,j}^{n} \sigma^{ij}\right). \tag{72}$$

A simple expression can be found for $\bar{n}$, provided the sum of all elements of the inverse correlation matrix is a *linear* function of n. As an example, consider noise with the exponential correlation function (*i.e.*, RC-filtered noise). The inverse covariance matrix in this case has already been referred to above. One finds that

$$E\left(\sum_{i,j}^{n} \sigma^{ij}\right) = [(\bar{n} - 2)(1 - \rho) + 2]/(1 + \rho), \tag{73}$$

from which it follows that for this example

$$\bar{n} = \frac{1+\rho}{1-\rho} \cdot \frac{L \log B + (1 - L) \log A}{(2aa_1 - a_1^2)/2} - \frac{2\rho}{1-\rho}. \tag{74}$$

When there is no correlation between observations ($\rho = 0$), (74) reduces to a result given by Wald. The second term in (74) represent the end effects and will

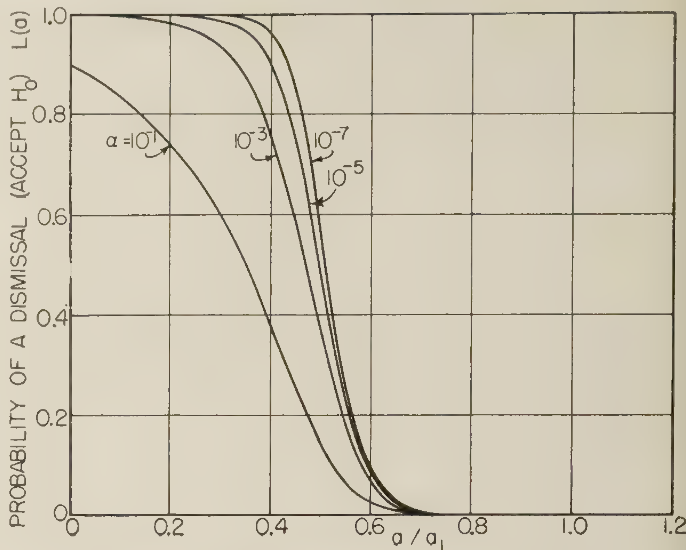


Fig. 5—OCF of the coherent detector; $\beta = 10^{-5}$.

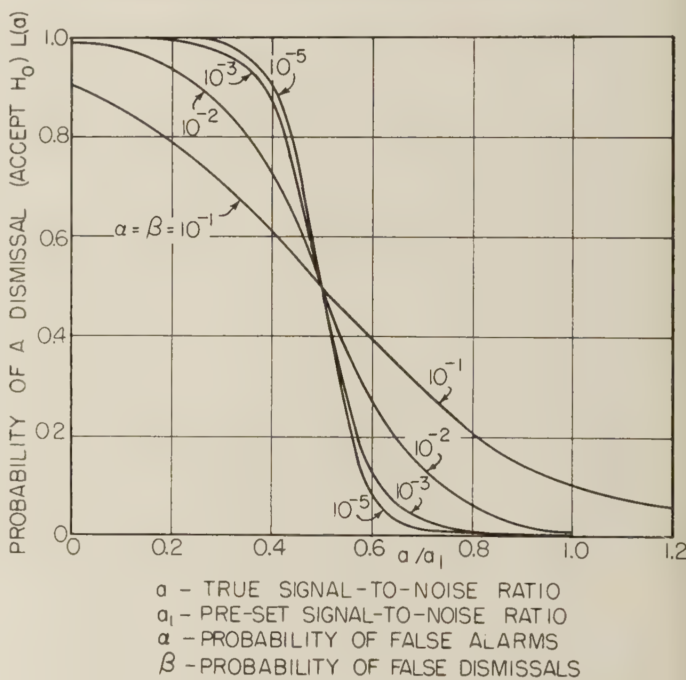


Fig. 6—OCF of the coherent detector; $\alpha = \beta$.

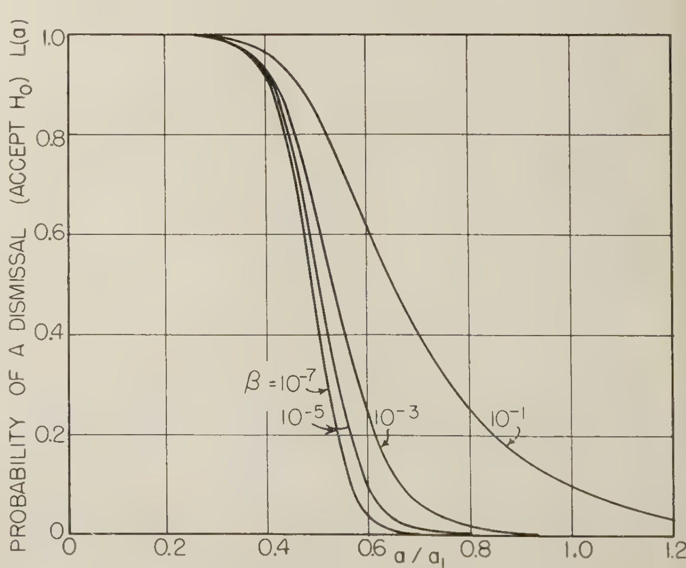


Fig. 7—OCF of the coherent detector; $\alpha = 10^{-5}$.

not usually be significant. If ρ is set equal to 1 (complete correlation), $\bar{n}$ becomes infinite. The stronger is the dependence between successive observations, the longer the test will last, which is consistent with the fact that little new information is supplied by each additional observation.

A series of curves illustrating (74), with $\rho = 0$, is shown in Fig. 8. Notice the typical peak of the ASN occurring between $a = 0$ and $a = a_1$. The peak becomes more pronounced as β decreases. It is also interesting to observe that for sequential tests, the behavior near $a = 0$ depends almost entirely on β, and the behavior near $a = a_1$, almost entirely on α.

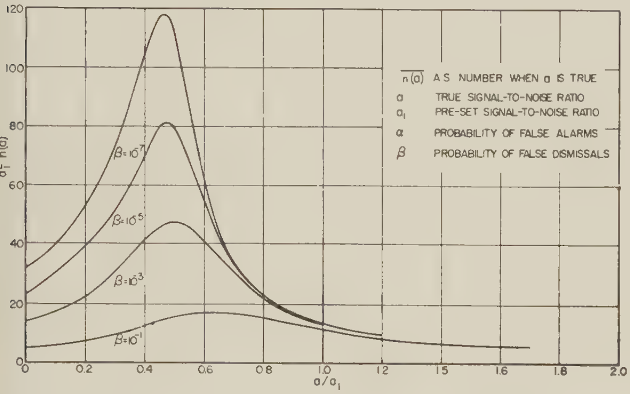


Fig. 8—Average sample numbers ($\times a_1^4$) vs a/a_1; coherent detection. $\alpha = 10^{-3}$; ($\rho = 0$).

Detection of Gaussian Random Signals in Normal Noise

Consider the case of Gaussian noise with a zero mean and a mean square value ψ. Let the signal also have Gaussian statistics with a zero mean but with a mean square value A^2. Assume that both have the same spectral shape (are passed through the same filter) and that the normalized correlation matrix of observations is given by $[\sigma_{ij}]$. Denote the sequence of observations by $X_1, X_2, \cdots$ so that $\sigma_{ij} = \overline{X_i X_j}/(\psi + A^2)$. The purpose of the test is to distinguish between the null hypothesis H_0 that $A = 0$ (no signal) and the alternative hypothesis H_1 that $A = A_1$ (signal A_1 present). Statistically this is the test for variance of a Gaussian distribution with a known mean, in which observations are not necessarily independent.

The first step in construction of such a test is to find the logarithm of the probability ratio. Let us introduce the normalization $a = A/\psi^{\frac{1}{2}}$. The distribution function of a sample of m observations is then

$$W_m(\mathbf{X}; a) = [2\pi\psi(1 + a^2)]^{-m/2}[\det(\sigma_{ij})]^{-1/2} \cdot \exp\left(-\frac{1}{2}\sum_{i,j}\frac{\sigma^{ij}}{1 - a^2}\frac{X_i X_j}{\psi}\right). \qquad (75)$$

The logarithm of the probability ratio, $Z_m = \log[W_m(\mathbf{X}; a_1)/W_m(\mathbf{X}; 0)]$, is therefore given by

$$Z_m = -\frac{1}{2} m \log(1 + a_1^2) + \frac{a_1^2}{1 + a_1^2}\sum_{i,j}\sigma^{ij}X_i X_j/\psi. \qquad (76)$$

The form of the detector is now in principle known. The actual instrumentation of a detector so defined must, of course, depend on the matrix $[\sigma^{ij}]$. For the example of RC-filtered "white" noise, which we have already once used as an illustration, the matrix $[\sigma^{ij}]$ is specified in the preceding section. Substituting into (76) and introducing a difference operator $\Delta^2 X_i = X_{i+1} - 2X_i + X_{i-1}$, we find for this particular example that

$$Z_m = -\frac{1}{2} m \log(1 + a_1^2) + \frac{1}{2\psi}\frac{a_1^2}{1 + a_1^2}\left[\sum_{i=2}^{m-1}\left(\frac{1-\rho}{1+\rho}X_i^2 - \frac{\rho}{1-\rho^2}X_i\Delta^2 X_{i-1}\right) + (1-\rho^2)^{-1}(X_1^2 + X_m^2 - \rho X_1 X_m - \rho X_m X_{m-1})\right]. \qquad (77)$$

If the observations are uncorrelated ($\rho = 0$), we get simply

$$Z_m = -\frac{1}{2} m \log(1 + a_1^2) + \frac{1}{2\psi}\frac{a_1^2}{1 + a_1^2}\sum_{i=1}^{m} X_i^2. \qquad (78)$$

Once again, the Operating Characteristic Function $L(a) = (A^h - 1)/(A^h - B^h)$ involves the solution of the parametric equation

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[W_m(\mathbf{X}; a_1)/W_m(\mathbf{X}; 0)]^{h(a, a_1, m)} \cdot W_m(\mathbf{X}; a)\, d\mathbf{X} = 1. \qquad (79)$$

Substituting from (75) and carrying out the integration, we find that the condition on h is in our case

$$h a_1^2(1 + a^2)/(1 + a_1^2) = 1 - (1 + a_1^2)^{-h}. \qquad (80)$$

This expression is identical with the relation for h which one would obtain if observations were uncorrelated [9], and it is likewise independent of sample size.

We conclude that for a sequential probability ratio detector and Gaussian noise and signal, the probability of which decision will terminate the detection process is a function of the actual (a) and the preset (a_1) signal-to-noise ratios only. It does not depend on the frequency spectrum nor on the spacing between observations.

In order to study the ASN, we consider $\overline{Z}_n$, the expected value of Z_n. Using (76), we get

$$\overline{Z}_n = -\frac{1}{2}\bar{n}\log(1 + a_1^2) + \frac{1}{2}\frac{a_1^2}{1 + a_1^2}E\left(\sum_{i,j}^{n}\sigma^{ij}X_i X_j/\psi\right). \qquad (81)$$

We take the statistical average of each term of the sum. But $\overline{X_i X_j} = \sigma_{ij}\,\psi(1 + a^2)$. We also have the relation

$$\sum_{i=1}^{n}\sigma^{ij}\sigma_{ij}/\psi = (1 + a^2)n, \qquad (\sigma_{ij} = \sigma_{ji}). \qquad (82)$$

Hence it follows that

$$\overline{Z}_n = -\tfrac{1}{2}\bar{n}[\log(1 + a_1^2) - a_1^2(1 + a^2)/(1 + a_1^2)]. \qquad (83)$$

Using the basic expression (70), we find finally that

$$\bar{n} = \frac{L\log\beta + (1 - L)\log A}{[-\log(1 + a_1^2) + a_1^2(1 + a^2)/(1 + a_1^2)]/2}. \qquad (84)$$

The relation (84) does not depend on the matrix $[\sigma^{ij}]$ and is the same as if the observations were uncorrelated [9]. We have already determined that the same holds true for (80). We conclude that for a normal random signal and normal noise with identical frequency spectra, the

number of observations required depends only on the signal-to-noise ratios a and a_1, and is independent of the correlation between observations. Thus there is no dependence either on the actual spectrum nor on the spacing between observations.

This result implies that by selecting points as close together in time as we wish, we can make the test last, *on the average*, an *arbitrarily short time*. The physical significance of this unusual result is closely connected with the assumption that a precise correlation matrix of observations is known and used in detection. A similar notion that arbitrarily strong tests can be performed over arbitrarily short time intervals was discovered by W. C. Fox [15] in connection with a study of fixed-size sample plans. We stress that our conclusion applies only under the conditions of this particular example; the process must have Gaussian statistics and the matrix $[\sigma^{ij}]$ must be known exactly and used in the construction of Z_m.

Truncated Tests

There are two reasons why a standard sequential procedure may be unsatisfactory. These are: 1) even if signal is present, an individual test may last longer than can be tolerated, 2) the average length of the test, when signal appears below the pre-set level, becomes extremely large, if the probabilities of error α and β are chosen to be very small. In some situations it may become virtually necessary to interrupt the standard procedure and resolve between the alternate courses of action, there and then.

If this is done, we speak of a *truncated sequential test.* Tests of this type were originally proposed by Wald [9]. The new rules of procedure can be fixed as follows: Carry out the regular sequential test until either a decision is made or stage N of the test is reached. At stage N, if no decision has been reached under the sequential rules of operation, accept the hypothesis H_0, if $Z_N \leq 0$ or accept the hypothesis H_1 if $Z_N \geq 0$. Under this new rule the test must terminate in at most N stages. Truncation is then a compromise between an entirely sequential test and a classical, fixed-sample test. It is an attempt to reconcile the good features of both of them: the sequential feature of examining observations as they accumulate and the classical feature of guaranteeing that the tolerances will be met with a specified sample size.

A truncated test can last no longer than N observations. The Average Sample Number $\bar{n}_T$ of a truncated test will differ from the ASN $\bar{n}$ of an untruncated test, even although both use the same bounds A and B. The difference can be expressed in terms of the probability density $P(n; a)$ of the size n of an untruncated test. We get [22]

$$\bar{n} - \bar{n}_T = \int_N^\infty (n - N)P(n; a)\, dn. \tag{85}$$

Since this integral is always positive it follows that a truncated test will take, on the average, a smaller sample than an untruncated test. This does not contradict the optimum property of sequential detectors, because a truncated process has no longer probabilities of error (α, β), but has some modified strength (α', β').

Wald [9] has established upper bounds on α' and β' for the case of normal distribution of z and independent observations. His upper bounds are

$$\alpha' \leq \alpha + [G(\nu_0') - G(\nu_0)] \tag{86}$$

and

$$\beta' \leq \beta + [G(\nu_1) - G(\nu_1')] \tag{87}$$

where
$$G(t) = (2\pi)^{-1} \int_{-\infty}^{t} \exp(-x^2/2)\, dx, \tag{88}$$

$$\nu_k = -N\overline{z(a_k)}/[N\sigma_z^2(a_k)]^{1/2} \tag{89}$$

and

$$\nu_k' = [\log K - N\overline{z(a_k)}]/[N\sigma_z^2(a_k)]^{1/2} \tag{90}$$

$$K = A \quad \text{if} \quad k = 0 \quad \text{and} \quad = B \quad \text{if} \quad k = 1. \tag{91}$$

Wald's upper bounds turn out to be appreciably above the actual probabilities of error. These probabilities, rather than just the actual bounds, can be expressed by introducing $p(n \mid d_k; a_j)$, the conditional distribution of n under the restriction that the terminal decision is to accept H_k when a_j is true ($k = 0, 1; j = 0, 1$). Notice that

$$P(n; a) = Lp(n \mid d_0; a) + (1 - L)p(n \mid d_1; a) \tag{92}$$

in which all the three distribution functions refer to an untruncated test. It can be shown [22] that

$$\alpha' = \alpha \int_0^N p(n \mid d_1; 0)\, dn + \gamma_0[G(\nu_0) - G(\nu_0')] \int_N^\infty P(n; 0)\, dn \tag{93}$$

and

$$\beta' = \beta \int_0^N p(n \mid d_0; a_1)\, dn + \gamma_{a_1}[G(\nu_1') - G(\nu_1)] \int_N^\infty P(n; a_1)\, dn, \tag{94}$$

in which

$$\gamma_a = \left[\int_{\log B}^{\log A} W(Z_N; a)\, dZ_N\right]^{-1} \tag{95}$$

and $W(Z_N; a)$ is the probability density of Z_N. A comparison of Wald's upper bounds in (86) and (87) with the expressions in (93) and (94) reveals that the latter differ by taking into account the probability that the truncation stage is reached. It can be verified; *e.g.*, that if $\alpha = \beta$, and $N = \bar{n}$ the relation (86) gives $\alpha' \leq (\alpha + \frac{1}{2})$, while (93) gives $\alpha' = \frac{1}{2}(\alpha + \frac{1}{2})$, so that the upper bound can be as much as twice the actual value.

Comparison of Sequential with Nonsequential Optimum Detectors

In comparing sequential with fixed-sample size (conventional) optimum detectors, we must remember that the sample size of a sequential test is a random variable whose expected value depends on the true signal-to-noise ratio. Thus when we compare sample sizes of two tests

we must specify what signal-to-noise ratio is assumed actually present.

Consider a sequential and a conventional test of the same strength (α, β), and with the same preset signal-to-noise ratio a_1. *The Average Percentage Saving S* is defined by

$$S(a) = 100[1 - \overline{n(a)}/n_f]\%, \tag{96}$$

in which n_f is the sample size of the conventional test. For a normal distribution of the logarithm of probability ratio (for all practical purposes this is the case of small α and β) and independent observations, it turns out [9, 22] that

$$S(a_1) = 100\{1 + g(\beta, \alpha)/[\Theta^{-1}(1 - 2\alpha) + \Theta^{-1}(1 - 2\beta)]^2\}\% \tag{97}$$

in which

$$\Theta(u) = (2/\sqrt{\pi}) \int_0^u dy \exp(-y^2).$$

Another interesting way to compare two detectors is to require that the strength (α, β) and the average sample sizes be identical $n_f = \overline{n(a_s)}$ and then compare the minimum detectable signals (say, a_c and a_s). The saving in the minimum detectable signal power is

$$S = 100(1 - a_s^2/a_k^2)\%. \tag{98}$$

For a coherent detector it turns out that the functions given by (96) and (97) are identical. Typical curves are shown in Fig. 9.

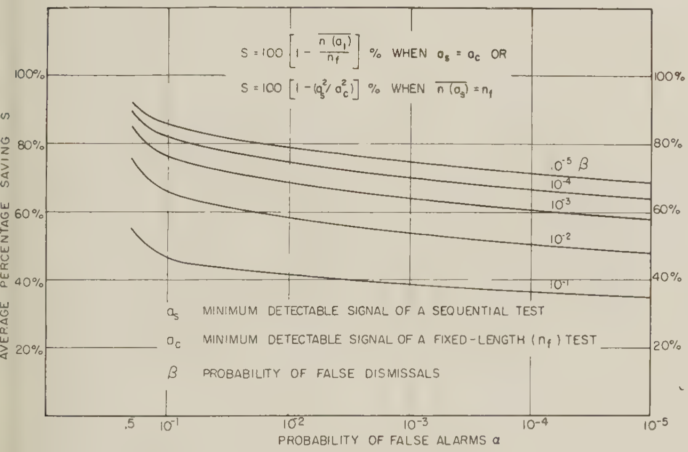


Fig. 9—Average percentage saving; coherent detection.

A comparison of the minimum detectable signals for the incoherent, sequential and "ideal" [8] modes of operation ($\alpha = \beta$; weak signals) is made in Fig. 10. Under the circumstances shown, a sequential observer detects signals 2–3 *db* weaker.

For a complete analysis of sequential detectors the distribution function of n is needed. Some studies of the sequential distribution functions have been carried out [9, 22]. One general observation may be made: sequential detection can be, *on the average*, carried out faster than conventional detection; however, this occurs at the expense of the detection time becoming a random variable. Thus, large savings in the sample size are associated with a large variance[3]) and there is an appreciable probability that a particular run will exceed the desired average length. Typically, with $\alpha = \beta = 10^{-5}$, the dispersion will be about 50 per cent of the average sample size.

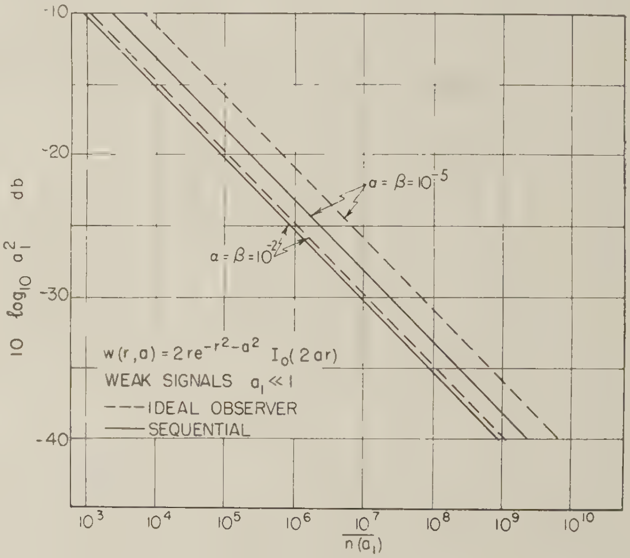


Fig. 10—Minimum detectable signal; comparison of observers. a_1^2 vs $\overline{n(a_1)}$.

List of Major Symbols

A	upper boundary of a sequential test
ASN	Average Sample Number
a	true value of the signal-to-noise amplitude ratio
$a_0 = 0$	signal-to-noise ratio associated with hypothesis H_0
a_1	signal-to-noise ratio associated with hypothesis H_1.
B	lower boundary of a sequential test
d_k	terminal decision to accept hypothesis H_k $(k = 0, 1)$
$E[n(\theta)]$	expected value of n when θ is true
H_0	null hypothesis: $a = 0$
H_1	alternative hypothesis: $a = a_1$
$L(a)$	probability of accepting H_0
m	a stage of the sequential detection process
n	the terminal stage of a sequential detection process
$\overline{n(\theta)}$	alternative notation for $E[n(\theta)]$.
N	truncation stage
OCF	Operating Characteristic Function
$P(n; a)$	distribution function of n when a is true
$p(n \mid d_k; a)$	conditional distribution function of n when a is true under the restriction d_k
T	time from the start of a sequential detection process (cf. m)
$\mathfrak{T}$	time from start to termination of a sequential detection process (cf. n)
$\bar{\mathfrak{T}}$	Average Sample Length (cf. $\bar{n}$)

[3] See remarks at the end of the first section of Part II, concerning the necessity of having a large variance if the sequential test is to have a different performance from a fixed sample test.

$w_m(\mathbf{x}; a)$	m-dimensional probability distribution of sample $x_1, x_2, \cdots, x_m$
$\mathbf{x}$	sample composed of observations $x_1, x_2, \cdots, x_m$
Z_m	logarithm of probability ratio of m observations
Z_T	logarithm of probability ratio over interval T
$\overline{Z}$	average of the logarithm of probability ratio
z	logarithm of probability ratio for a single observation
α	probability of a false alarm
α'	probability of a false alarm for a truncated test
β	probability of a false dismissal
β'	probability of a false dismissal for a truncated test
λ_m	probability ratio at the mth observation

Bibliography

[1] Marcum, J. I., *A Statistical Theory of Target Detection by Pulsed Radar*, Rand Corp., RM-754, December 1, 1951; ibid., "Mathematical Appendix," RM-753, July 1, 1948.

[2] Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," D.Sc. Thesis, Massachusetts Institute of Technology, January, 1951.

[3] Schwartz, J., "A Statistical Approach to the Automatic Search Problem," Ph.D. Thesis, Harvard University, June 1951.

[4] Slattery, T. G., "The Detection of Sine Wave in Presence of Noise by the Use of Nonlinear Filter," *Proceedings of the IRE*, Vol. 40 (October, 1952), p. 1232.

[5] Middleton, D., *Statistical Criteria for the Detection of Pulsed Carriers in Noise*, Air Force Cambridge Res. Cent. Rep., November, 1952.

[6] Reich, E. and P. Swerling, "The Detection of a Sinewave in Gaussian Noise," *Journal of Applied Physics*, Vol. 24 (March 1953), p. 289.

[7] Middleton, D. "Statistical Criteria for Detection of Pulsed Carriers in Noise: I, II," *Journal of Applied Physics*, Vol. 24 (April 1953), pp. 371–378; pp. 379–391.

[8] Siegert, A. J. F., in J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, Radiation Laboratory Series (McGraw-Hill, 1950), Vol. 24, Ch. 7.

[9] Wald, A., *Sequential Analysis*. New York, John Wiley and Sons, Inc., 1947.

[10] Wald, A. and J. Wolfowitz, "Optimum Character of the Sequential Probability Ratio Test," *Annals of Mathematical Statistics*, Vol. 19, p. 326 (1948).

[11] The Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data: Applications*, Columbia University Press, New York, 1945.

[12] Stein, C., "A Note on Cumulative Sums," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 498–499.

[13] Girshick, M. A., "Contributions to the Theory of Sequential Analysis I, II and III," *Annals of Mathematical Statistics*, Vol. 17 (1946).

[14] Barnard, G. A., "Sequential Tests in Industrial Statistics," *Supplement of the Journal of the Royal Statistical Society*, Vol. 8, No. 1 (1946).

[15] Fox, W. C., *Signal Detectability: A Unified Description of Statistical Methods Employing Fixed and Sequential Observation Processes*, Electronic Defense Group, Univ. of Michigan, Tech. Rep. No. 19, December, 1953.

[16] Peterson, W. W., T. G. Birdsall, and W. C. Fox, "The Theory of Signal Detectability," 1954 Symposium on Information Theory, *Transactions of the IRE* (*PGIT*), Vol. 4 (September, 1954), pp. 179–182.

[17] Harrington, J. V., *An Analysis of Detection of Repeated Signals in Noise by Binary Integration*, MIT Lincoln Lab., Tech. Rep. No. 13, August 14, 1952, and *Transactions of the IRE* (*PGIT*), Vol. 1 (March, 1955), pp. 1–9.

[18] Reed, I. S., *Analysis of Signal Detection by Sequential Observer*, MIT Lincoln Lab., Tech. Rep. No. 20, March 12, 1953.

[19] Reed, I. S. and G. P. Dineen, MIT Lincoln Lab., personal communication.

[20] Blasbalg, H., Theory of Sequential Filtering and Its Application to Signal Detection and Classification, Johns Hopkins University, Radiation Lab., Tech. Rep. No. AF-8, October 18, 1954.

[21] Bussgang, J. J., "Sequential Detection of Signals in Noise," Ph.D. Thesis, Harvard University, 1955.

[22] Bussgang, J. J., and D. Middleton, "Sequential Detection of Signals in Noise," Harvard Cruft Lab., Tech. Rep. No. 175, 1955.

[23] Van Meter, D. and D. Middleton, "Modern Statistical Approaches to Reception in Communication Theory," *Transactions of the IRE* (*PGIT*), Vol. 4 (September, 1954); p. 119.

[24] Middleton, D. and D. Van Meter, "Detection and Extraction of Signals in Noise from the Viewpoint of Statistical Decision Theory," Soc. for Ind. and Appl. Math., Vol. 4, No. 4 (Dec., 1955) and Vol. 5, No. 1 (March, 1956).

[25] Wald, A., *Statistical Decision Functions*. New York, John Wiley and Sons, Inc., 1950.

[26] Lehmann, E. L., "On Families of Admissible Tests," *Annals of Mathematical Statistics*, Vol. 18, 1947.

[27] Stein, C. M., "On Sequences of Experiments," abstract in *Annals of Mathematical Statistics*, Vol. 19 (1948).

[28] Blackwell, D. and M. A. Girshick, *Theory of Games and Statistical Decisions*. New York, John Wiley Sons, Inc., 1954.

[29] Mood, A., *Introduction to the Theory of Statistics*. New York, McGraw-Hill, 1950.

[30] Darling, D. A. and A. J. F. Siegert, "The First Passage Problem for a Continuous Markov Process," *Annals of Mathematical Statistics*, Vol. 24 (December 1953), pp. 624–639.

[31] Kemperman, F., "The General One-Dimensional Random Walk with Absorbing Barriers with Applications to Sequential Analysis," University of Amsterdam, Thèse de Doctorat, 1950.

[32] Rice, S. O., *Mathematical Analysis of Random Noise*. Bell Telephone Systems, Monograph B-1589, 1944–1945.

[33] Woodward, P. M., *Probability and Information Theory, with Applications to Radar*. New York, McGraw-Hill, 1953.

[34] Middleton, D., The Statistical Theory of Detection I: Optimum Detection of Signals in Noise, MIT Lincoln Lab., Tech. Rep. No. 35, November, 1953.

[35] Stein, S. and J. E. Storer, "Generating a Gaussian Sample," to be published in the 1955 Wescon Convention Record.

On Binary Channels and Their Cascades

RICHARD A. SILVERMAN†

Summary—**A detailed analysis of the general binary channel is given, with special reference to capacity (both separately and in cascade), input and output symbol distributions, and probability of error. The infinite number of binary channels with the same capacity lie on double-branched equicapacity lines. Of the channels on the lower branch of a given equicapacity line, the symmetric channel has the smallest probability of error and the largest capacity in cascade, unless the capacity is small, in which case the asymmetric channel (with one noiseless symbol) has the smallest probability of error and the largest capacity in cascade. By simply reversing the designation of the output (or input) symbols, we can decrease the probability of error of any channel on the upper branch of the equicapacity line and increase the capacity in cascade of any asymmetric channel on the upper branch.**

In a binary channel neither symbol should be transmitted with a probability lying outside the interval [1/e, 1 − (1/e)] if capacity is to be achieved. The maximally asymmetric input symbol distributions are approached by certain low-capacity channels. For these channels, redundancy coding permits an appreciable fraction of capacity in cascade if sufficient delay can be tolerated.

Capacity and Symbol Distributions

Discussion of the binary channel is usually confined to the symmetric case, where each of the transmitted digits is similarly perturbed by the noise. However, many interesting features of binary channels are concealed if only *symmetric* channels are considered. Accordingly, this paper will be devoted to a detailed study of the arbitrary binary channel.

Let the channel be characterized by the transition-probability matrix

$$C = \begin{Vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{Vmatrix}, \qquad 0 \le \alpha, \beta \le 1, \tag{1}$$

where α is the probability that a zero be received as a zero, β the probability that a one be received as a zero, etc. We shall use the symbol C for both the channel and its matrix, but no confusion will arise. Computations are simplified by defining (after Muroga[1]) an auxiliary vector

$$\vec{X} = \begin{Vmatrix} X_0 \\ X_1 \end{Vmatrix}.$$

which solves the equation

$$C\vec{X} = -\vec{H}.$$

Here $\vec{H}$ is the row-entropy vector of the channel C; *i.e.*,

$$\vec{H} = \begin{Vmatrix} H(\alpha) \\ H(\beta) \end{Vmatrix}.$$

where $H(x)$ is the entropy function

$$H(x) = -x \log x - (1-x) \log (1-x).$$

(All logarithms are to the base 2 unless otherwise indicated; as another notation for 2^x, we shall write $\exp_2 x$.) Muroga[1] has shown that the capacity $c(C)$ of the channel C can be written in terms of the components of the vector $\vec{X}$ as

$$c(C) = \log (2^{X_0} + 2^{X_1}). \tag{2}$$

The transmitted and received symbol distributions are also simply related to $\vec{X}$. Let $\vec{P}$ be a vector representing the transmitted symbol distribution which achieves capacity, *i.e.*,

$$\vec{P} = \begin{Vmatrix} P_0 \\ 1 - P_0 \end{Vmatrix},$$

where P_0 is the probability with which zeros should be chosen if capacity (maximum rate) is to be achieved. Let $\vec{P}'$ be a vector representing the corresponding received symbol distribution, *i.e.*,

$$\vec{P}' = \begin{Vmatrix} P_0' \\ 1 - P_0' \end{Vmatrix},$$

where P_0' is the probability that a zero will be received if the transmitted symbol distribution achieves capacity. The vectors $\vec{P}$ and $\vec{P}'$ are related by the equation

$$\vec{P}' = \tilde{C}\vec{P},$$

where $\tilde{C}$ denotes the transpose of the matrix C. In terms of the auxiliary vector $\vec{X}$, Muroga finds that

$$P_0 = \frac{2^{-c}}{\det (C)} \det \begin{Vmatrix} 2^{X_0} & 2^{X_1} \\ \beta & 1-\beta \end{Vmatrix}, \tag{3}$$

and that

$$P_0' = 2^{X_0 - c}. \tag{4}$$

Our task is to express the quantities (2), (3), and (4) in terms of the parameters α and β of the binary channel (1). After some algebraic manipulation we find that

$$c(\alpha, \beta) = \frac{-\beta H(\alpha) + \alpha H(\beta)}{\beta - \alpha} + \log \left[1 + \exp_2 \left(\frac{H(\alpha) - H(\beta)}{\beta - \alpha} \right) \right], \tag{5}$$

$$P_0(\alpha, \beta) = \beta(\beta - \alpha)^{-1} - (\beta - \alpha)^{-1} \left[1 + \exp_2 \left(\frac{H(\beta) - H(\alpha)}{\beta - \alpha} \right) \right]^{-1}, \tag{6}$$

$$P_0'(\alpha, \beta) = \left[1 + \exp_2 \left(\frac{H(\beta) - H(\alpha)}{\beta - \alpha} \right) \right]^{-1}. \tag{7}$$

† Formerly Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.; now Institute of Mathematical Sciences, Div. of Electromagnetic Res., New York Univ., N. Y.

[1] S. Muroga, "On the capacity of a discrete channel. I. *J. Phys. Soc. Japan*, vol. 8, pp. 484–494; July–August, 1953.

Each of the quantities (5), (6), and (7) depends on the channel parameters α and β: (5) gives the capacity of C, (6) the probability with which zeros should be chosen at the transmitter if capacity is to be achieved, and (7) the probability of a zero appearing at the receiver if zeros are chosen at the transmitter in accordance with (6).

Each of the functions $c(\alpha, \beta)$, $P_0(\alpha, \beta)$, and $P_0'(\alpha, \beta)$ defines a surface over the unit square $0 \leq \alpha, \beta \leq 1$. A study of the expressions (5), (6), and (7) reveals the following symmetries:

$$c(\alpha, \beta) = c(\beta, \alpha) = c(1-\alpha, 1-\beta) = c(1-\beta, 1-\alpha), \quad (8)$$

$$P_0(\alpha, \beta) = P_0(1-\alpha, 1-\beta) = 1 - P_0(\beta, \alpha) = 1 - P_0(1-\beta, 1-\alpha), \quad (9)$$

$$P_0'(\alpha, \beta) = P_0'(\beta, \alpha) = 1 - P_0'(1-\alpha, 1-\beta) = 1 - P_0'(1-\beta, 1-\alpha). \quad (10)$$

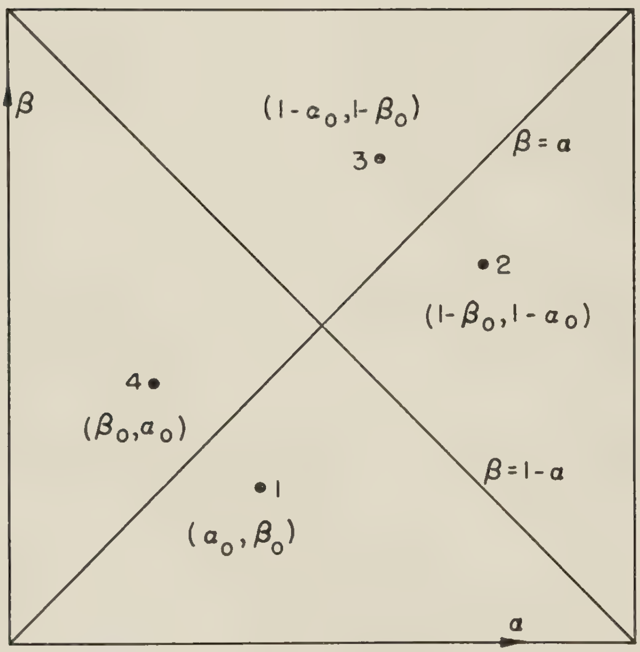


Fig. 1—Illustrating the symmetries of $c(\alpha, \beta)$, $P_o(\alpha, \beta)$, etc.

These symmetries are illustrated in Fig. 1, where a point (α_0, β_0) and its reflections in the $\beta = \alpha$ and $\beta = 1 - \alpha$ lines are shown. Eq. (8) shows that the capacity has the same value at any four such symmetrically placed points (for a reason to be discussed shortly), (9) that P_0 has the same value at points 1 and 3 and one minus that value at points 2 and 4, and (10) that P_0' has the same value at points 1 and 4 and one minus that value at points 2 and 3.

Fig. 2 shows lines of constant capacity (equicapacity lines). Along the line $\beta = \alpha$ the capacity vanishes, corresponding to the vanishing of det (C). The line $\beta = 1 - \alpha$ is the locus of symmetric channels, and along this line (5) reduces to the familiar expression

$$c(\alpha, \alpha) = 1 - H(\alpha).$$

Along the lines $\alpha = 0$, $\alpha = 1$, $\beta = 0$, and $\beta = 1$, (5) reduces to especially simple expressions. For example,

$$c(\alpha, 0) = \log [1 + \exp_2 (-H(\alpha)/\alpha)], \qquad 0 \leq \beta \leq 1.$$

The slope of the curves $c(\alpha, 0)$ and $c(0, \beta)$ at the point $\alpha = \beta = 0$ is $\log e/e$. It is clear from Fig. 2 that there are an infinite number of binary channels with the same capacity. This is to be expected since *two* parameters are required to uniquely specify a binary channel.

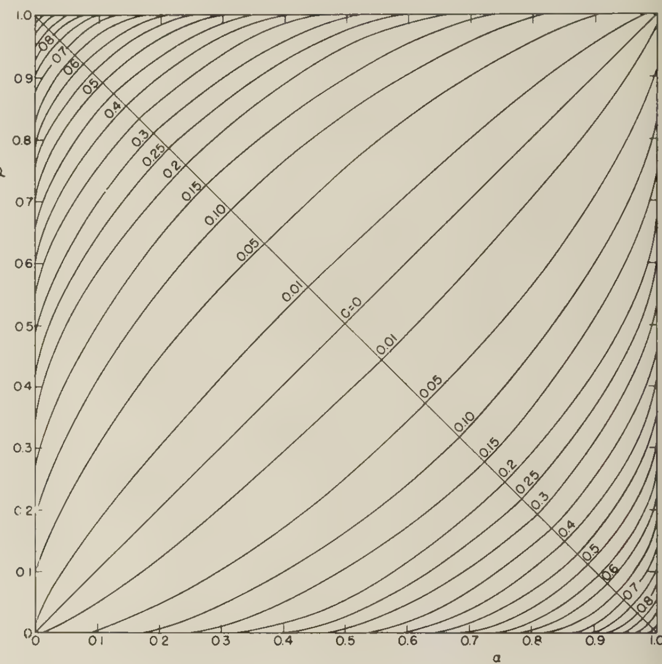


Fig. 2—Lines of constant $c(\alpha, \beta)$.

The fact that the four channels $C(\alpha, \beta)$, $C(\beta, \alpha)$, $C(1-\alpha, 1-\beta)$, and $C(1-\beta, 1-\alpha)$ have the same capacity, which produces two symmetrically placed branches of each equicapacity curve is easily explained. Clearly it is a matter of indifference which input (or output) symbol we choose to call a zero and which we choose to call a one. Reversing the designation of the input symbols corresponds to premultiplication by the noiseless matrix

$$I = \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix},$$

and maps the channel $C(\alpha, \beta)$ into the channel $C(\beta, \alpha)$. Reversing the designation of the output symbols corresponds to postmultiplication by the matrix I and maps the channel $C(\alpha, \beta)$ into the channel $C(1-\alpha, 1-\beta)$. Reversing the designation of *both* the input and output symbols corresponds to premultiplication *and* postmultiplication by the matrix I, and maps the matrix $C(\alpha, \beta)$ into the matrix $C(1-\beta, 1-\alpha)$. As (9) and (10) show, there are properties that, unlike capacity, are not invariant under all these mappings.

We have just seen that from a given point on an equicapacity line at least three other points can be

reached by multiplying the channel matrix by another channel matrix. That there are no more such points is an immediate consequence of a theorem proved by DeSoer[2] to the effect that the capacity of two channels in cascade is less than the capacity of either unless one is a noiseless channel (*i.e.*, the unit matrix or one of its permutations) or unless one is a completely noisy (zero-capacity) channel.[3,4] (The reader is reminded that connecting two (or more) channels in cascade corresponds to multiplying the corresponding matrices.) Our statement follows from the fact that there are only two noiseless binary channels, namely I and the unit matrix.

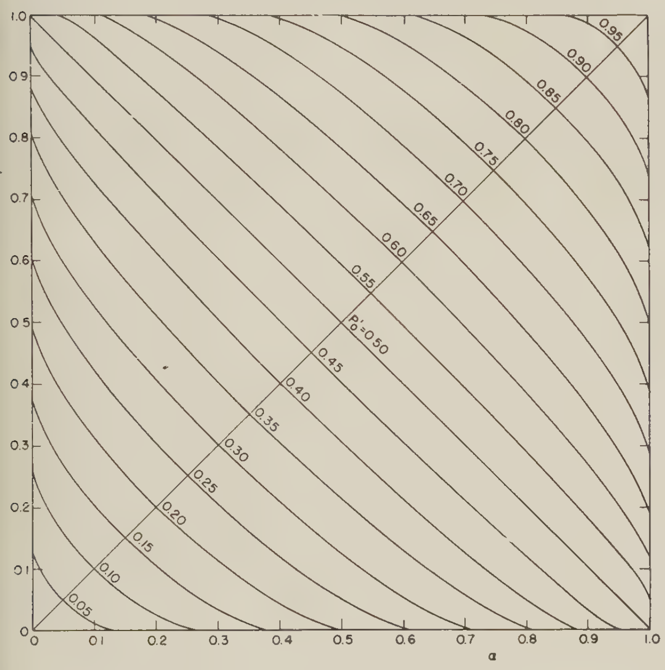


Fig. 3—Lines of constant $P'_o(\alpha, \beta)$.

Fig. 3 shows lines of constant $P'_0(\alpha, \beta)$, the probability of receiving a zero if the input symbol distribution achieves capacity. Along the line $\beta = 1 - \alpha$, the locus of symmetric channels, $P'_0(\alpha, \beta)$ has the familiar value $\frac{1}{2}$. Along the zero-capacity line $\beta = \alpha$, $P'_0(\alpha, \beta)$ has the limiting value α, although (7) is indeterminate for $\beta = \alpha$. That is

$$\lim_{\epsilon\to 0} P'_0(\alpha + \epsilon, \alpha) = \lim_{\epsilon\to 0} P'_0(\alpha, \alpha + \epsilon) = \alpha.$$

Fig. 4 shows lines of constant $P_0(\alpha, \beta)$, the probability with which zeros should be transmitted to achieve capacity. The surface is saddle-shaped with the saddle point at $\alpha = \beta = \frac{1}{2}$. Along the symmetric channel line $\beta = 1 - \alpha$, $P_0(\alpha, \beta)$ has the familiar value $\frac{1}{2}$. Along the zero-capacity line $\beta = \alpha$, $P_0(\alpha, \beta)$ has the limiting value $\frac{1}{2}$, although (6) is indeterminate for $\beta = \alpha$. That is,

$$\lim_{\epsilon\to 0} P_0(\alpha + \epsilon, \alpha) = \lim_{\epsilon\to 0} P_0(\alpha, \alpha + \epsilon) = \tfrac{1}{2}.$$

The behavior of $P_0(\alpha, \beta)$ at the corners $\alpha = \beta = 0$ and $\alpha = \beta = 1$ is sufficiently remarkable to warrant special discussion.

Suppose we approach the point $\alpha = \beta = 0$ along the line $\alpha = \epsilon$, $\beta = r\epsilon$, where $0 \leq r < \infty$; *i.e.*, along any line between the positive α-axis ($r = 0$) and the positive β-axis ($r = \infty$). Then $\lim_{\epsilon\to 0} P(\epsilon, r\epsilon)$ takes on all values between $1/e$ and $1 - (1/e)$, depending on the value of r, provided that for the value $r = 1$ (for which the single limit is indeterminate) we take the double limit $\lim_{\epsilon\to 0} \lim_{\epsilon'\to 0} P_0(\epsilon, \epsilon + \epsilon')$. For example,

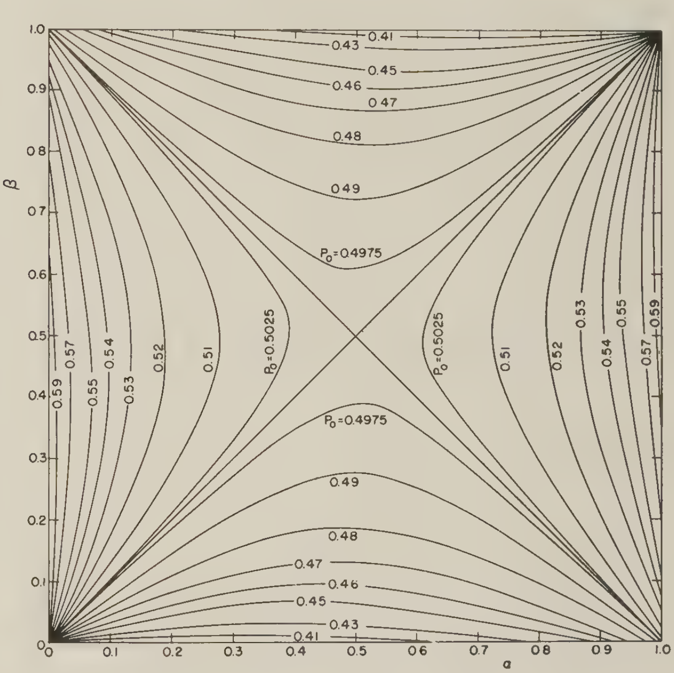


Fig. 4—Lines of constant $P_o(\alpha, \beta)$.

$$\lim_{\epsilon\to 0} P_0(\epsilon, 0) = \lim_{\epsilon\to 0} [\epsilon(1 + \exp_2 (H(\epsilon)/\epsilon)]^{-1}$$

$$= \lim_{\epsilon\to 0} \left[e + \left(1 - \frac{e}{2}\right)\epsilon\right]^{-1} = 1/e,$$

$$\lim_{\epsilon\to 0} P_0(0, \epsilon) = 1 - \lim_{\epsilon\to 0} P_0(\epsilon, 0) = 1 - (1/e),$$

$$\lim_{\epsilon\to 0} P_0(\epsilon, \tfrac{1}{2}\epsilon) = \lim_{\epsilon\to 0} - 1 + [(e/4) + (\epsilon/2) - (3e\epsilon/16)]^{-1}$$

$$= (4/e) - 1,$$

$$\lim_{\epsilon\to 0}\lim_{\epsilon'\to 0} P_0(\epsilon, \epsilon + \epsilon') = \lim_{\epsilon\to 0}\lim_{\epsilon'\to 0}$$

$$1 + \frac{\epsilon(1 - \epsilon)}{2 \log e} H''(\epsilon) + \epsilon 0(\epsilon') = \frac{1}{2}.$$

[2] C. A. DeSoer, "Communication through channels in cascade," Sc.D. Thesis, January, 1953, Dept. of Elect. Eng., M.I.T.

[3] There is also the intermediate case where the channel matrix is reducible and one of the submatrices is completely noisy, *e.g.*, the channel with matrix

$$\begin{Vmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

This channel (cited by Shannon in reference 4) is effectively the unit matrix, since the symbols corresponding to the first and second rows produce indistinguishable effects at the receiver.

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949, pp. 44–45.

In evaluating the limits we have made free use of the expressions

$$H(x) = x \log (e/x) - (x^2/2) \log e + 0(x^3),$$

$$H'(x) = (d/dx)H(x) = \log (1 - x)/x,$$

$$-H''(x) = -(d^2/dx^2)H(x) = x^{-1}(1 - x)^{-1} \log e.$$

We see that the point $\alpha = \beta = 0$ (and its image $\alpha = \beta = 1$ in the line $\beta = 1 - \alpha$) is a point of discontinuity of the function $P_0(\alpha, \beta)$, whose limiting behavior there depends on the direction of approach in a sort of "spiral staircase" fashion.

The maximum value of $P_0(\alpha, \beta)$ is the limiting value $1 - (1/e)$ obtained when we approach the point $\alpha = \beta = 0$ along the positive β-axis, and the minimum value of $P_0(\alpha, \beta)$ is the limiting value $1/e$ obtained when we approach the point $\alpha = \beta = 0$ along the positive α-axis. (There are two corresponding limits at the point $\alpha = \beta = 1$.) Of course, the channel capacity is zero in both limits, so that there is no channel with positive capacity whose input symbol distribution is as asymmetric as $\vec{P} = [1/e, 1 - (1/e)]$ or $\vec{P} = [1 - (1/e), 1/e]$. However, there are low-capacity channels whose input symbol distributions are arbitrarily close to these maximally asymmetric ones. These low-capacity channels will be discussed further below.

We see that in a binary channel neither transmitted symbol can be selected with a probability lying outside the interval $[1/e, 1 - (1/e)]$ if capacity is to be achieved. If we are compelled to send digits from a more asymmetric distribution (as we may well be), the possibility of signaling at capacity is precluded from the start. Intuitively, this means that in a binary channel no decrease in equivocation obtained by skewing the input symbol distribution can justify making the source entropy less than $H(1/e)$, at least if obtaining maximum rate (capacity) is the objective.

For channels with larger alphabets it may be quite proper to choose one or more transmitted symbols with probability less than $1/e$, or indeed to suppress one or more transmitted symbols. Thus, for example, in the ternary channel

$$\begin{Vmatrix} \alpha & 1 - \alpha & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 - \alpha & \alpha \end{Vmatrix},$$

capacity cannot be achieved if the symbol corresponding to the second row is transmitted, unless $\alpha < \alpha_0$, where $\alpha_0 \sim 0.64$ is the solution of the equation

$$\log \alpha = -\alpha.$$

Muroga gives many other examples of the need for suppressing possible transmitted symbols in his basic paper[1]. He was the first to point out the need of taking special care that P_0 does not become negative in capacity calculations.

The four channels which in the zero-capacity limit achieve one of the maximally asymmetric input distributions $\vec{P} = [1/e, 1 - (1/e)]$ or $\vec{P} = [1 - (1/e), 1/e]$ have matrices

$$\begin{Vmatrix} \epsilon & 1 - \epsilon \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1 - \epsilon & \epsilon \\ 1 & 0 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 \\ \epsilon & 1 - \epsilon \end{Vmatrix}, \quad \begin{Vmatrix} 1 & 0 \\ 1 - \epsilon & \epsilon \end{Vmatrix}. \tag{11}$$

We shall refer to these channels as "ϵ-channels". To the first order in ϵ, they all have capacity

$$c = \frac{\log e}{e} \epsilon \sim 0.53\epsilon \text{ bits.} \tag{12}$$

Introducing the abbreviation $k = (2 - e)/e \sim -0.27$, we find that the input symbol distribution which achieves capacity for the *first* and *second* of these channels is

$$\vec{P} = \begin{Vmatrix} (e + k\epsilon)^{-1} \\ 1 - (e + k\epsilon)^{-1} \end{Vmatrix}, \tag{13}$$

whereas for the *third* and *fourth* it is

$$\vec{P} = \begin{Vmatrix} 1 - (e + k\epsilon)^{-1} \\ (e + k\epsilon)^{-1} \end{Vmatrix}. \tag{14}$$

The corresponding output symbol distributions are

$$\vec{P}' = \begin{Vmatrix} \epsilon(e + k\epsilon)^{-1} \\ 1 - \epsilon(e + k\epsilon)^{-1} \end{Vmatrix} \tag{15}$$

for the *first* and *third* channels, and

$$\vec{P}' = \begin{Vmatrix} 1 - \epsilon(e + k\epsilon)^{-1} \\ \epsilon(e + k\epsilon)^{-1} \end{Vmatrix} \tag{16}$$

for the *second* and *fourth* channels. Eqs. (13) through (16) are accurate only to the first order in ϵ.

Probability of Error

We have seen that there are infinitely many binary channels with the same capacity. It is natural to ask whether there are contexts in which any of these channels with the same capacity is to be preferred to the others. Two questions that we might ask are: 1) Which of the channels with the same capacity has the smallest probability of error (in a received digit), and 2) Which has the largest end-to-end capacity if its output terminals are connected to the input terminals of an identical channel?[5] We answer the first question in this section and defer discussion of the second question until the next section.

The probability of error (at capacity) is given by the expression

$$P_e(\alpha, \beta) = \beta + (1 - \alpha - \beta)P_0(\alpha, \beta), \tag{17}$$

[5] This question arises naturally if we consider building up a cascade of repeaters, using a given binary channel as a unit.

where $P_0(\alpha, \beta)$ is the probability of a transmitted zero as given by (6). It is easily verified that $P_e(\alpha, \beta)$ has the following symmetries

$$P_e(\alpha, \beta) = P_e(1-\beta, 1-\alpha) = 1 - P_e(\beta, \alpha)$$
$$= 1 - P_e(1-\alpha, 1-\beta). \tag{18}$$

In deriving (18) free use has been made of the symmetries of $P_0(\alpha, \beta)$ as given by (9). Referring to Fig. 1, we see that $P_e(\alpha, \beta)$ has the same value at the points 1 and 2 and one minus that value at the points 3 and 4. (Note that none of the functions $c(\alpha, \beta)$, $P_0(\alpha, \beta)$, $P_0'(\alpha, \beta)$, and $P_e(\alpha, \beta)$ has the same symmetries.)

Fig. 5 shows lines of constant $P_e(\alpha, \beta)$. Along the symmetric channel line $\beta = 1 - \alpha$, $P_e(\alpha, 1-\alpha)$ has the familiar value $1 - \alpha$. Along the zero-capacity line $\beta = \alpha$, $P_e(\alpha, \alpha)$ has the limiting value $\frac{1}{2}$. At the point $\alpha = 0$, $\beta = 1$ corresponding to the channel matrix I, $P_e = 1$, whereas at the point $\alpha = 1$, $\beta = 0$ corresponding to the unit matrix, $P_e = 0$.

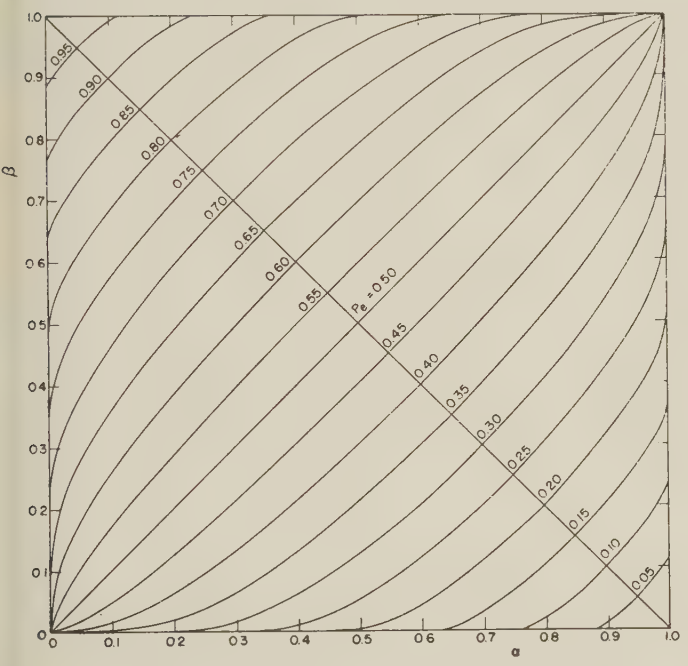


Fig. 5—Lines of constant $P_e(\alpha, \beta)$.

We have already noted that $P_0(\alpha, \beta)$ has discontinuities at the corners $\alpha = \beta = 0$ and $\alpha = \beta = 1$, and indeed that any value between $1/e$ and $1 - (1/e)$ can be obtained by approaching these discontinuities along the proper directions. Eq. (17) shows that $P_e(\alpha, \beta)$ shares these discontinuities. For since

$$\lim_{\alpha,\beta\to 0} P_e(\alpha, \beta) = \lim_{\alpha,\beta\to 0} P_0(\alpha, \beta),$$

the limiting behavior of $P_e(\alpha, \beta)$ at the two discontinuities is identical with that of $P_0(\alpha, \beta)$, however different the over-all appearance of the two sets of curves. This fact is apparent from Fig. 5. In particular, it follows that the curve $P_e(\alpha, \beta) = 1/e$ (not shown) must come into the points $\alpha = \beta = 0$ and $\alpha = \beta = 1$ with zero slope.

As we have seen, the portion of the equicapacity curve lying in the triangular region $\beta \leq \alpha$, $\beta \leq 1 - \alpha$ generates the rest of the equicapacity curve under the mappings corresponding to reversing the designation of the input symbols or output symbols or both. Eq. (18) shows that the probability of error for channels on the upper branch of the equicapacity curve (above the line $\beta = \alpha$) is greater than for channels on the lower branch (below the line $\beta = \alpha$), and indeed is greater than $\frac{1}{2}$. However, a value of P_e greater than $\frac{1}{2}$ is artificial, for if communication is through such a channel, the receiver can obtain information at the same rate and with probability of error one minus that value merely by reversing the designation of the received symbols. (The transmitter need not be informed of this reversal, for (9) shows that the input symbol distribution remains the same in the reversed channel.) Thus our problem reduces to finding which of the channels on the portion of the equicapacity curve lying in the triangular region $\beta \leq \alpha$, $\beta \leq 1 - \alpha$ has the smallest probability of error.

The question is immediately answered if we superimpose the curves of Figs. 2 and 5. We find that a symmetric channel with a given capacity has a smaller probability of error than an asymmetric channel with the same capacity, *unless the channels have very low capacity.* In the latter case it is easily verified that the symmetric channel

$$\left\| \begin{matrix} 1/2 + (\epsilon/2e)^{1/2} & 1/2 - (\epsilon/2e)^{1/2} \\ 1/2 - (\epsilon/2e)^{1/2} & 1/2 + (\epsilon/2e)^{1/2} \end{matrix} \right\|$$

has the same capacity as the asymmetric ϵ channel

$$\left\| \begin{matrix} \epsilon & 1 - \epsilon \\ 0 & 1 \end{matrix} \right\|,$$

namely

$$\frac{\log e}{e} \epsilon \text{ bits.}$$

Using (13) and (17), we find that the probability of error for the asymmetric channel is

$$(P_e)_{\text{asym.}} = (1-\epsilon)/(e + k\epsilon) \qquad [k = (2-e)/e],$$

whereas that of the symmetric channel is obviously

$$(P_e)_{\text{sym.}} = 1/2 - (\epsilon/2e)^{1/2}.$$

For small ϵ, it is apparent that

$$(P_e)_{\text{asym.}} < (P_e)_{\text{sym.}},$$

as asserted. Indeed

$$\lim_{\epsilon\to 0} (P_e)_{\text{asym.}} = 1/e, \tag{19}$$

whereas

$$\lim_{\epsilon\to 0} (P_e)_{\text{sym.}} = \tfrac{1}{2}. \tag{20}$$

Cascaded Channels

We turn to a discussion of the second of the questions raised at the beginning of the preceding section: Which of the channels with the same capacity has the largest end-to-end capacity if its output terminals are connected to the input terminals of an identical channel?

First we remind the reader (see the first section) that

$$c(C^2) < c(C),$$

i.e., capacity is a decreasing function of the number of cascaded stages, unless C is one of the noiseless channels

$$\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}, \qquad \begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix},$$

or one of the zero-capacity channels

$$\begin{Vmatrix} a & 1-a \\ a & 1-a \end{Vmatrix}. \tag{21}$$

The reader should further note that, unless C is one of the two noiseless channels,

$$\lim_{n\to\infty} c(C^n) = 0.$$

For then C is irreducible, and by a well-known theorem on Markov chains[6] $\lim_{n\to\infty} C^n$ has the form (21), and consequently

$$\lim_{n\to\infty} c(C^n) = c(\lim_{n\to\infty} C^n) = 0.$$

We begin by squaring the matrix $C(\alpha, \beta)$, obtaining

$$C^2 = \begin{Vmatrix} A(\alpha,\beta) & 1 - A(\alpha,\beta) \\ B(\alpha,\beta) & 1 - B(\alpha,\beta) \end{Vmatrix},$$

where

$$A(\alpha, \beta) = \alpha^2 + (1-\alpha)\beta,$$
$$B(\alpha, \beta) = \alpha\beta + (1-\beta)\beta.$$

Thus to every channel (α, β) on a given equicapacity curve corresponds a squared channel (cascaded with itself with matrix elements $A(\alpha, \beta)$ and $B(\alpha, \beta)$. The functions $A(\alpha, \beta)$ and $B(\alpha, \beta)$ exhibit the following symmetries:

$$\begin{aligned} A(\alpha, \beta) &= 1 - B(1-\beta, 1-\alpha), \\ B(\alpha, \beta) &= 1 - A(1-\beta, 1-\alpha), \end{aligned} \tag{22}$$

and

$$\begin{aligned} A(\beta, \alpha) &= 1 - B(1-\alpha, 1-\beta), \\ B(\beta, \alpha) &= 1 - A(1-\alpha, 1-\beta). \end{aligned} \tag{23}$$

However, there is no simple relation between $A(\alpha, \beta)$ and $A(\beta, \alpha)$, or between $B(\alpha, \beta)$ and $B(\beta, \alpha)$ [unless $\beta = 1 - \alpha$], so that two quite different curves are produced by squaring the matrices corresponding to a given equicapacity line, one originating from the upper branch of the equicapacity curve (above the line $\beta = \alpha$), and the other originating from the lower branch (below the line $\beta = \alpha$). The portion of the (A, B) curve which originates from the upper branch of the equicapacity curve has no intercepts on the α- and β-axes; we shall call it the *upper branch* of the (A, B) curve. The portion of the (A, B) curve which originates from the lower branch of the equicapacity curve has intercepts on the α- and β-axes; we shall call it the *lower branch* of the (A, B) curve. Eqs. (22) and (23) show that both branches of the (A, B) curve are symmetric in the line $\beta = 1 - \alpha$. Moreover, since

$$A(\alpha, 1-\alpha) = A(1-\alpha, \alpha),$$
$$B(\alpha, 1-\alpha) = B(1-\alpha, \alpha),$$

the two branches contact on the line $\beta = 1 - \alpha$.

These facts are illustrated by Fig. 6, which shows three (A, B) curves, those generated by squaring the channels with capacities 0.1, 0.4, and 0.7. The (α, β) values corresponding to these channels were read off the corresponding equicapacity lines of Fig. 2. Note that all the (A, B) curves lie below the line $\beta = \alpha$. This is because $\beta > \alpha$ implies $B(\alpha, \beta) < A(\alpha, \beta)$, whereas $\beta < \alpha$ implies $B(\beta, \alpha) < A(\beta, \alpha)$.

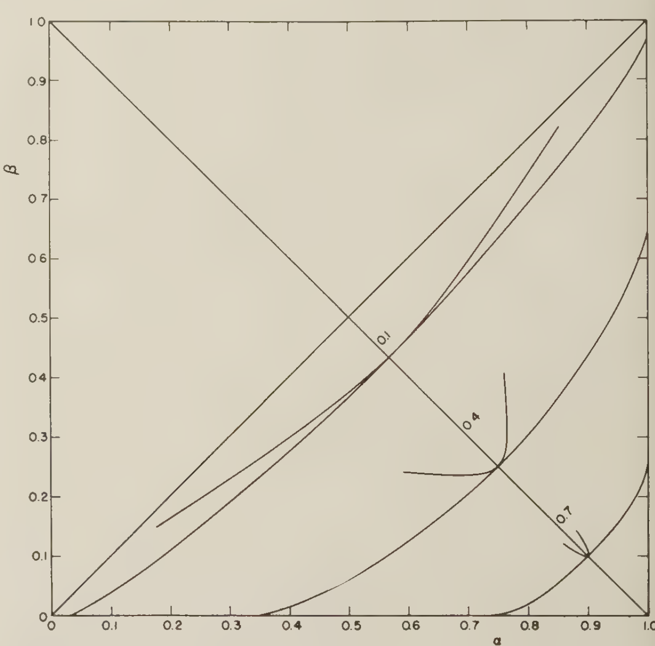


Fig. 6—(A, B) curves corresponding to the channels of capacities 0.1, 0.4, and 0.7.

Of all the channels with the same capacity, the two corresponding to the end-points of the upper branch of the (A, B) curve have the smallest capacity under cascade. Moreover, any channel on the upper branch of the (A, B) curve has lower capacity than its images (under multiplication by I) on the lower branch. However, we can avoid the low capacity in cascade exhibited by these channels by an extremely simple intermediate station behavior, namely by crossing the connections between the outputs

[6] W. Feller, "Probability Theory and Its Applications," vol. 1 New York, John Wiley and Sons, 1950. Reference is made to Theorem 2, p. 325.

of one channel and the inputs of the next. In this way we arrive on the lower branch of the (A, B) curve, which has higher capacity in cascade. (It will be recalled that in the preceding section we avoided probabilities of error greater than $\frac{1}{2}$ by exactly the same expedient.) We shall comment on the significance of this intermediate station behavior below.

The problem thus reduces to finding the channel on the lower branch of the (A, B) curve with the highest capacity. As in the preceding section, we resort to a superposition of curves, this time superimposing the curves of Figs. 2 and 6. We find that a symmetric channel with a given capacity has a higher capacity under cascade than an asymmetric channel with the same capacity, *unless the channels have very low capacity.* As an example of this exceptional behavior at low capacity, we cite again the two channels

$$\begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1/2 + (\epsilon/2e)^{1/2} & 1/2 - (\epsilon/2e)^{1/2} \\ 1/2 - (\epsilon/2e)^{1/2} & 1/2 + (\epsilon/2e)^{1/2} \end{Vmatrix}, \tag{24}$$

which both have capacity of $(\epsilon \log e)/e$ bits. The squares of the matrices (24) are

$$\begin{Vmatrix} \epsilon^2 & 1-\epsilon^2 \\ 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1/2 + \epsilon/e & 1/2 - \epsilon/e \\ 1/2 - \epsilon/e & 1/2 + \epsilon/e \end{Vmatrix},$$

respectively. The corresponding capacities are

$$c[C^2_{\text{asym.}}] = \frac{\log e}{e} \epsilon^2 \text{ bits}, \tag{25}$$

and

$$c[C^2_{\text{sym.}}] = \frac{2 \log e}{e^2} \epsilon^2 \text{ bits} \tag{26}$$

Thus

$$c[C^2_{\text{asym.}}] > c[C^2_{\text{sym.}}],$$

as asserted.[7]

We can now answer both questions posed at the beginning of the preceding section as follows: Of all the binary channels on the lower branch of a given equicapacity line, the symmetric channel has the smallest probability of error and the largest capacity in cascade, *unless the capacity is small,* in which case the asymmetric channel (with one noiseless symbol, *i.e.*, $\beta = 0$) has the smallest probability of error and the largest capacity in cascade. Continuity requires that there be a small range of values of the capacity for which asymmetric channels with $\beta \neq 0$ are superior in these two respects, but we have made no detailed study of this intermediate case. There is no point in trying to establish a preference among the channels on the upper branch of the equicapacity line, since by reversing the designation of the output (or input) symbols we arrive at a channel with a smaller probability of error and a larger capacity in cascade. (An exception to this statement in the symmetric case is noted below.)

[7] Eqs. (19), (20), (25), and (26) suggest the conjecture that the capacity in cascade of channels with the same *low* capacity is inversely proportional to their probabilities of error (see also (32) in the Appendix).

A particularly striking example of the difference between the capacity in cascade of a channel on the upper branch of the equicapacity line and its images on the lower branch is afforded by comparing the two ϵ-channels

$$C_\epsilon = \begin{Vmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{Vmatrix}, \qquad C'_\epsilon = C_\epsilon I = \begin{Vmatrix} 1-\epsilon & \epsilon \\ 1 & 0 \end{Vmatrix}.$$

C_ϵ is on the lower branch of the equicapacity line (with capacity $(\epsilon \log e)/e$ bits), and C'_ϵ, obtained by reversing the designation of the output terminals of C_ϵ, is on the upper branch. The squares of C_ϵ and C'_ϵ are

$$C^2_\epsilon = \begin{Vmatrix} \epsilon^2 & 1-\epsilon^2 \\ 0 & 1 \end{Vmatrix}$$

and

$$C'^2_\epsilon = \begin{Vmatrix} 1-\epsilon+\epsilon^2 & \epsilon-\epsilon^2 \\ 1-\epsilon & \epsilon \end{Vmatrix}.$$

The corresponding capacities are

$$c[C^2_\epsilon] = \frac{\log e}{e} \epsilon^2 \text{ bits}$$

and

$$c[C'^2_\epsilon] = \frac{\log e}{8} \epsilon^3 \text{ bits},$$

so that, in this extreme case, the capacity of one channel is an order of magnitude less than that of the other.

DeSoer[2] has emphasized the importance of proper intermediate station behavior in maximizing the end-to-end capacity of a cascade of channels. In particular, he compares the capacity of a cascade of continuous channels perturbed by white Gaussian noise with that of a cascade of PCM channels with the same signal-to-noise ratio. It is assumed that in the cascade of continuous channels the intermediate stations retransmit the received waveform without change, whereas in the cascade of PCM channels requantization occurs at each intermediate station. Although the continuous channel has a higher capacity than the PCM channel, the PCM channels deteriorate less in cascade. Thus, for some cascade length depending on the signal-to-noise ratio, the cascade of PCM channels has a larger end-to-end capacity than the cascade of continuous channels.

We have an even simpler example of how proper intermediate station behavior can preserve the end-to-end capacity of cascaded channels. For, in the example just given, if the output symbols of the C'_ϵ channel are reversed at the intermediate station, we have

$$c[C'_\epsilon \, I \, C'_\epsilon] = c[C^2_\epsilon] \gg c[C'^2_\epsilon].$$

On the other hand, the capacity of a cascade of *symmetric* channels is completely insensitive to whether the identity of the symbols is preserved or reversed at the intermediate

stations. Graphically, this is a consequence of the point of contact of the two branches of the (A, B) curve on the symmetric channel line $\beta = 1 - \alpha$ (See Fig. 6).

Suppose we regard the behavior of the intermediate station as a detection scheme. Then preserving the designation of the output symbols of the preceding channel at the intermediate station is minimum probability of error detection[8] if the channel lies in the square region $\alpha \geq \frac{1}{2}, \beta \leq \frac{1}{2}$, and reversing the designation of the output symbols is minimum probability of error detection if the channel lies in the square region $\alpha \leq \frac{1}{2}, \beta \geq \frac{1}{2}$. If the channel lies in the square region $\alpha, \beta \leq \frac{1}{2}$, then minimum probability of error detection requires that both zeros and ones be changed to ones, and information-destroying mapping. Similarly, if the channel lies in the square region $\alpha, \beta \geq \frac{1}{2}$, then minimum probability of error detection requires that both zeros and ones be changed to zeros. Thus it is apparent that maximum rate in cascade and minimum probability of error detection at intermediate stations are not always compatible. (DeSoer[2] gives a complicated example that illustrates this fact.) If information-destroying mappings are precluded, as they must be if maximum rate is the objective, we conclude that the larger capacity in cascade is given by minimum probability of error detection, which requires that the identity of the symbols be preserved at the intermediate station if the channel lies below the line $\beta = \alpha$, but reversed if the channel lies above the line $\beta = \alpha$.

The proofs of the statements of the preceding paragraph are left to the reader. It is merely necessary to examine the expressions for the probabilities of error of each of the four delayless detection schemes and ascertain which is smallest in each square region.

Appendix

Redundancy Coding in the ϵ-Channel

In our discussion of cascaded channels in the last section we considered only delayless operation of the intermediate station. If sufficient intermediate station delay is allowed, it follows from Shannon's second coding theorem that the end-to-end capacity of a cascade of identical channels can be made arbitrarily close to the common capacity of the separate channels. Studies of probability of error and rate as a function of delay are still in progress,[9] and it is perhaps too early to apply the theory to cascaded channels. However, the ϵ-channel is susceptible to a simple type of *redundancy* coding, which is effective just because of its low capacity. This redundancy coding, although not ideal in the sense of achieving capacity with a vanishingly small probability of error, nonetheless achieves a rate which is an appreciable fraction of capacity with a small probability of error. Moreover, it serves to illustrate how delay can be exchanged for enhanced rate in a cascade of channels.[10]

In the redundancy coding to which we refer, each transmitted digit is repeated r times, and the receiver decides whether a zero or a one was sent by examining sequences of r digits. More specifically, let the channel have matrix

$$C_\epsilon = \begin{Vmatrix} \epsilon & 1 - \epsilon \\ 0 & 1 \end{Vmatrix},$$

with capacity $(\log e)/e\ \epsilon$ bits. Have the receiver examine the output in blocks of r symbols (properly synchronized with the transmitter) and decode a block of r ones as a one and a block of r digits with a zero at *any* position as a zero. In other words, the symbols 0 and 1 are mapped into the sequences $00 \cdots 0$ (r times) and $11 \cdots 1$ (r times) at the transmitter, and the events S and F are mapped into 0 and 1 at the receiver, where S designates the appearance of a zero in a block of r digits, and F the nonappearance of a zero in a block of r digits.[11] Transmission can then be regarded as taking place in an equivalent channel $C(r)$ with matrix

$$C(r) = \begin{Vmatrix} 1 - (1 - \epsilon)^r & (1 - \epsilon)^r \\ 0 & 1 \end{Vmatrix}.$$

The capacity of $C(r)$ is

$$c[C(r)] = \log \left[1 + \exp_2 \left(\frac{-H((1 - \epsilon)^r)}{1 - (1 - \epsilon)^r} \right) \right],$$

and its probability of error is

$$P_e(r) = (1 - \epsilon)^r P_0(r),$$

where by $P_0(r)$ we mean the probability that the sequence $00 \cdots 0$ (r times) should be transmitted if the capacity $c[C(r)]$ is to be achieved. $P_0(r)$ is not the same as P_0 for the ϵ-channel, as given by (13) of the first section.

Suppose that each of the transmitted symbols is repeated $r = n/\epsilon$ times. Then, since $(1 - \epsilon)^{n/\epsilon} \sim e^{-n}$ for small ϵ, the equivalent channel becomes

$$C(n) = \begin{Vmatrix} 1 - e^{-n} & e^{-n} \\ 0 & 1 \end{Vmatrix},$$

with capacity

$$c[C(n)] = \log \left[1 + \exp_2 \left(\frac{-H(1 - e^{-n})}{1 - e^{-n}} \right) \right], \tag{27}$$

input symbol distribution

$$P_0(n) = (1 - e^{-n})^{-1} \left[1 + \exp_2 \left(\frac{H(1 - e^{-n})}{1 - e^{-n}} \right) \right]^{-1}, \tag{28}$$

and probability of error

$$P_e(n) = e^{-n} P_0(n). \tag{29}$$

[8] This is sometimes called maximum *a posteriori* probability detection or the ideal observer.

[9] Reference is made to recent work by C. E. Shannon and by P. Elias, presented at the March, 1955 National Convention of the IRE.

[10] Tentative studies of the effect of delay in cascaded channels have also been made by DeSoer[2].

[11] Of course, redundancy coding is also effective in a low capacity *symmetric* channel, but the analysis is more complicated, since now the events S and F refer to receiving more zeros than ones in a block of r received digits, and vice versa.

Of course, before comparing the capacity of the redundant channel with the capacity of the ϵ-channel, we must normalize (27) by dividing it by $r = n/\epsilon$, since only one information digit is transmitted every r units of time. Moreover, we must now accept a delay of r units. The interesting result (displayed in Table I at the end of this appendix) is that even when properly normalized, the redundancy code (which is, after all, a very simple code) gives rates which are an appreciable fraction of the capacity of the ϵ-channel, with a probability of error which becomes smaller as we tolerate more delay. (Unfortunately, the capacity goes to zero with the probability of error, which is not the case for ideal coding.) Note that as more redundancy is introduced, $C(n)$ becomes a better approximation to the unit matrix, and $P_0(n)$ approaches $\frac{1}{2}$.

TABLE I

n	$c[C(n)]$	$c[C(n)](\epsilon/n)$	$P_0(n)$	$P_e(n)$
1	0.436	0.436ϵ	0.413	0.152
2	0.707	0.353ϵ	0.448	0.061
3	0.858	0.286ϵ	0.472	0.024
4	0.934	0.234ϵ	0.487	0.0089
5	0.971	0.194ϵ	0.493	0.0033
6	0.987	0.165ϵ	0.497	0.0012
7	0.994	0.142ϵ	0.499	0.00045
8	0.998	0.125ϵ	0.500	0.00017
9	0.999	0.111ϵ	0.500	0.00006
10	1.000	0.100ϵ	0.500	0.00002

(For no redundancy: $c[C_\epsilon] = 0.531\epsilon$, $P_0 = 0.368$, $P_e = 0.368$)

Illustrating the redundancy-coded ϵ-channel. A table of the quantities given by (27), (28), and (29), corresponding to a redundancy $r = n/\epsilon$.

If $C(n)$ is cascaded N times, we raise the matrix $C(n)$ to the Nth power:

$$C^N(n) = \begin{Vmatrix} (1-e^{-n})^N & 1-(1-e^{-n})^N \\ 0 & 1 \end{Vmatrix}.$$

The capacity of the product matrix is still normalized by dividing it by the per-stage delay $r = n/\epsilon$, but the over-all delay that must now be tolerated is Nr. Assume that n is large enough so that e^{-n} is small and $c[C(n)] \sim 1.0$, and suppose that we agree to let $c[C^N(n)]$ fall off only to the value $k_1 < 1$. This amount of deterioration occurs when

$$(1-e^{-n})^N = f(k_1), \tag{30}$$

where $f(k_1)$ is the abscissa of the point of intersection of the equicapacity line k_1 with the α-axis (see Fig. 2). If we assume that N, the number of cascaded stages, is large (30) becomes

$$\exp(-Ne^{-n}) = f(k_1)$$

or

$$Ne^{-n} = -\log f(k_1) \equiv k_2 > 0.$$

Thus

$$N = k_2 e^n, \tag{31}$$

i.e., we can tolerate more channels in the cascade if we increase n, and consequently the error-proofing and delay per stage. Moreover, since

$$P_e(n) \sim \tfrac{1}{2}e^{-n},$$

(31) can be rewritten as

$$N \sim \frac{k_2}{2P_e(n)}. \tag{32}$$

Eq. (32) says that the amount of cascading permissible to within a given tolerated deterioration of the end-to-end capacity is inversely proportional to the probability of error per stage.[12]

If a noiseless feedback channel is available at the receiver, the rate can be increased by the simple expedient of having the receiver instruct the transmitter to begin a new run of r repetitions whenever a zero is received. For, since received zeros can only originate from transmitted zeros, it is a waste of channel space to continue repeating zeros for the rest of the run of r digits when a zero has already been received. In the limit of high redundancy this feedback procedure increases the rate by a factor approaching 2, because without feedback almost half the channel space is taken up by the needless repetition of zeros.

ACKNOWLEDGMENT

This work was done while the author held the position of Research Associate in the Department of Electrical Engineering of the Massachusetts Institute of Technology, and was supported by Industrial Fellowship funds of the Research Laboratory of Electronics.

The numerical calculations on which the figures and table are based were done by Mrs. Elgie G. Levin.

[12] If k_2 is very small, self-consistency may require n to be quite large, since N is assumed to be large (see (31)).

Minimum Energy Cost of An Observation

F. P. ADLER†

Summary—The minimum energy expenditure required in performing basic observations and measurements is analyzed. The energy cost, in ergs per binary unit (bit) of information, is found for three fundamental cases using idealized experimental procedures: 1) the determination of the presence (or absence) of an input signal on an indicating instrument, 2) the measurement of a time interval and 3) the measurement of a distance. The variation of energy cost with the reliability and accuracy of the experiment is determined; it is found that with a suitable procedure the minimum value of kT ln2 ergs per bit predicted by the Second Law (interpreted so as to include informational entropy) can be approached arbitrarily closely under conditions of small reliability and high accuracy. The present results are compared with those derivable from C. E. Shannon's equation for the capacity of a communication channel.

Introduction

THE FACT that an observation requires expenditure of energy and hence an increase in entropy has been discussed by several authors[1-13] who have also pointed out the resulting fallacy in the working of Maxwell's demon. In the present paper we are concerned with finding quantitatively the minimum energy cost of an observation or experiment, in ergs per bit of information gained. Our point of departure is Brillouin's article[10] in which he points out that in any experiment the sum of entropy minus information[14] can never decrease. The efficiency of an observation, which he defines as the ratio of information gain (ΔI_{th}) to entropy increase (ΔS), can therefore never exceed unity. Brillouin then illustrates this point using certain idealized basic measurement procedures. The following analysis is largely an extension and amplification, in quantitative terms, of the above paper. Although cost (c), rather than efficiency, will be considered here, the latter may readily be found from the relation

$$\eta = \frac{\Delta I_{th}}{\Delta S} = \frac{k\Delta I \ln 2}{E/T} = \frac{kT \ln 2}{E/\Delta I} = \frac{kT \ln 2}{c} \tag{1}$$

where ΔI_{th} is in thermodynamic units and ΔI is in bits; E denotes the added energy and T the absolute temperature. An immediate consequence of this conversion formula is seen to be that

$$c = kT \ln 2 \quad \text{ergs per bit} \tag{2}$$

represents the lower bound on information cost, corresponding to the ideal case of unity efficiency.

Observation of Single Oscillator

We shall first consider the case of taking a reading on an instrument such as a ballistic galvanometer to determine whether or not a certain signal is present. The indicating mechanism (needle and spring) constitutes an oscillatory system (natural frequency ν) subject to Brownian motion. Following Brillouin, we may therefore consider the observation of the needle position equivalent to determining the energy state of a harmonic resonator of frequency ν with quantized energy levels $nh\nu$. Now suppose that a reading corresponding to a particular energy level $mh\nu$ is observed. From the Boltzmann distribution law the probability that the resonator would attain this level due to thermal fluctuations only is

$$p_0(m) = Ke^{-mh\nu/kT} \tag{3}$$

where $K = 1 - \exp(-h\nu/kT)$ is a normalizing constant such that $\sum_{n=0}^{\infty} p_0(n) = 1$. If, on the other hand, an input signal has actually been applied to the resonator, then the probability of this level becomes

$$p_1(m) = Ke^{-(m-s)h\nu/kT} \qquad (m \geq s) \tag{4}$$

where s is the number of added signal quanta. Now the information gain, ΔI, will be the difference between I_0 and I_1, the information functions before and after the observation. Assuming the maximum a priori uncertainty, that is, signal and no signal equally likely, we have

$$I_0 = \log_2 2 = 1 \text{ bit} \tag{5}$$

while

$$I_1 = p_S \log_2 \frac{1}{p_S} + p_N \log_2 \frac{1}{p_N} \tag{6}$$

where p_S is the *a posteriori* probability that the observed energy level $mh\nu$ was indeed due to a signal and p_N that it was a spurious indication caused by thermal noise.

† Hughes Research and Development Labs., Culver City, Calif.

1 L. Szilard, "Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," *Z. Physik*, vol. 53, p. 840; March, 1929.

2 P. Demers, "Les démons de maxwell et le second principe de la thermodynamique," *Can. J. Research*, vol. 22, p. 27; March, 1944.

3 P. Demers, "Le second principe et la théorie des quanta," *Can. J. Research*, vol. 23, p. 47; May, 1945.

4 L. Brillouin, "Life, thermodynamics, and cybernetics," *Am. Scientist*, vol. 37, p. 554; October, 1949.

5 L. Brillouin, "Thermodynamics and information theory," *Am. Scientist*, vol. 38, p. 594; October, 1950.

6 L. Brillouin, "Maxwell's demon cannot operate: information and entropy," *J. Appl. Phys.*, vol. 22, p. 334; March, 1951.

7 L. Brillouin, "Physical entropy and information," *J. Appl. Phys.*, vol. 22, p. 338; March, 1951.

8 H. Jacobson, "The role of information theory in the inactivation of maxwell's demon," *Trans. N. Y. Acad. Sciences*, vol. 14, p. 6; November, 1951.

9 G. F. Cawsey, "Physical entropy and the entropy of information theory," Roy. Air. Estabnt. Tech. Note No. 169; February, 1952.

10 L. Brillouin, "The negentropy principle of information," *J. Appl. Phys.*, vol. 24, p. 1152; September, 1953.

11 L. Brillouin, "Negentropy and information in telecommunications, writing and reading," *J. Appl. Phys.*, vol. 25, p. 595; May, 1954.

12 L. Brillouin, "Information theory and uncertainty principle," *J. Appl. Phys.*, vol. 25. p. 887; July, 1954.

13 M. J. E. Golay, "Maxwell's demon must remain incommunicado," *J. Appl. Phys.*, vol. 25, p. 1062; August, 1954.

14 In order to be directly comparable with entropy, information must be measured in thermodynamic units; thus ΔI_{th} equals k times number of natural units ("nits") or k ln 2 times number of bits.

Bayes' theorem is now used to find p_S and p_N:

$$p_N = p(H_0/E) = \frac{p(E/H_0)p(H_0)}{p(E/H_0)p(H_0) + p(E/H_1)p(H_1)} \tag{7}$$

$$p_S = p(H_1/E) = \frac{p(E/H_1)p(H_1)}{p(E/H_0)p(H_0) + p(E/H_1)p(H_1)}$$

where

E .... observed event (m quanta); (8)

H_0 .... hypothesis that signal is not present;

H_1 .... hypothesis that signal is present;

$p(H_0), p(H_1)$. *a priori* probabilities of H_0 and H_1 (assumed equal to 1/2);

$p(E/H_0)$ .... probability of occurrence of E if H_0 is true = $p_0(m) = K \exp(-mh\nu/kT)$;

$p(E/H_1)$ .... probability of occurrence of E if H_1 is true = $p_1(m) = K \exp[-(m-s)h\nu/kT]$.

Putting in the values indicated yields

$$p_N = \frac{1}{1 + e^{sh\nu/kT}}$$

$$p_S = \frac{1}{1 + e^{-sh\nu/kT}}. \tag{9}$$

(9)[15]

The information gain is therefore

$$\Delta I(sh\nu/kT) = \Delta I(S)$$

$$= 1 - \frac{\log_2(1 + e^{-S})}{1 + e^{-S}} - \frac{\log_2(1 + e^{S})}{1 + e^{S}} \tag{10}$$

where $S \equiv sh\nu/kT$. Since it requires s signal quanta or SkT ergs to produce ΔI, the cost of the observation is seen to be

$$c_0 = \frac{SkT}{\Delta I(S)} \text{ ergs per bit.} \tag{11}$$

A plot of c_0 is shown in Fig. 1. The minimum cost is about 4.1 kT or 4.0×10^{-14} erg per bit at room temperature, corresponding to a maximum efficiency of 0.17. It occurs for a signal input of 2.57 kT. It is interesting to compare this value with the value of 4 kT considered by G. Ising[16] to be the minimum input required for a *reliable* signal indication; it is seen that high reliability must be paid for by increased cost and lowered efficiency.

Time Measurement

Here we shall take as our experimental model an idealized electronic timing device of the type suggested by Brillouin[10]: a number (N) of receivers are connected to a line from which a signal pulse is expected to come; the receivers are successively turned on, each for a time τ, and the occurrence in time of the pulse (relative to a zero-time reference pulse) is determined by noting which of the N receivers shows an output signal. The switching on of the receivers may be performed by means of timing pulses which change bias voltages on diodes or vacuum tube grids; this operation need not, in principle, require any energy expenditure. The energy of the signal pulse, however, is dissipated upon arrival at a receiver where it is absorbed and presumably recorded.

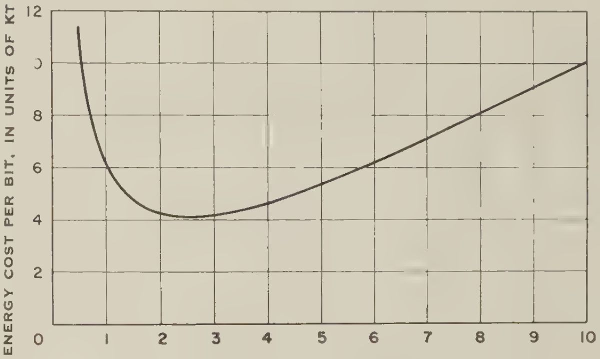


Fig. 1—Energy cost of observation on single resonator. Threshold energy is a measure of reliability of observation.

We shall consider an observational procedure in which a detector is considered as having received the signal pulse if its energy state equals or exceeds a certain threshold value, $E_t = n_t h\nu$[17]. If only one detector shows excitation above this level, the information gain is clearly

$$\Delta I = \log_2 N \tag{12}$$

assuming that initially all N receivers were equally likely to receive the signal pulse. The same result may of course also be obtained using the *a priori* distribution of the pulse occurrence time t, $p_0(t) = 1/N\tau$, and its *a posteriori* distribution, $p_1(t) = 1/\tau$, giving

$$\Delta I = \int_0^{N\tau} p_0 \log_2 \frac{1}{p_0}\,dt - \int_0^{\tau} p_1 \log_2 \frac{1}{p_1}\,dt = \log_2 N. \tag{13}$$

If m receivers, rather than just one, are observed to exceed n_t quanta, a unique measurement is not possible, but at least the possible values of the time interval have been narrowed down from N to m. Although this may perhaps be of somewhat questionable value to the experimenter, the information gain is clearly

$$\Delta I_m = \log_2 (N/m). \tag{14}$$

Now the probability that a receiver will equal or exceed the threshold level due to noise alone is

[15] It is of interest to note that because of the exponential nature of the Maxwell-Boltzmann distribution, p_N and p_S turn out to be actually independent of m. If this were not the case, the cost would need to be averaged with respect to m to find the mean cost over all possible observed states.

[16] G. Ising, "A natural limit for the sensibility of galvanometers," *Phil. Mag.*, vol. 1–7, p. 827; April, 1926.

[17] It may be thought that an increased amount of information could be obtained if instead of setting a threshold level, the actual amount of energy on each resonator were observed immediately subsequent to its active interval and inferences drawn from these values as to the likelihood of any detector having received a signal. As shown in the Appendix, however, such an experimental procedure turns out precisely equivalent to the one considered here.

$$p(n \geq n_t) = \sum_{n=n_t}^{\infty} p(n) = \sum_{n=n_t}^{\infty} Ke^{-nh\nu/kT} = e^{-A} \qquad (15)$$

where A denotes the normalized energy threshold $n_t h\nu/kT$. The probability of a total of m receivers giving an indication is therefore

$$p_m = \frac{(N-1)!}{(N-m)!(m-1)!} e^{-(m-1)A}(1-e^{-A})^{N-m}$$
$$= \binom{N-1}{m-1} e^{-(m-1)A}(1-e^{-A})^{N-m}$$
$$\left(\sum_{m=1}^{N} p_m = 1\right) \qquad (16)$$

which is the product of the probability that $(m-1)$ receivers will indicate a signal and the probability that the remaining $(N-m)$ will not, multiplied by the number of ways of realizing this situation with the $(N-1)$ receivers which do not receive the signal pulse. The *average* information gain is then given by

$$\Delta I = \sum_{m=1}^{N} p_m \log_2 (N/m). \qquad (17)$$

The necessary signal energy is determined by the requirement that the signal pulse should always produce an indication, that is, equal or exceed the threshold energy E_t; clearly, a signal energy equal to E_t will produce this effect, even if the receiving resonator should happen to be in a zero quantum state. The minimum cost of the time observation is therefore

$$c_t = \frac{E_t}{\Delta I}$$
$$= E_t\left[\sum_{m=1}^{N}\binom{N-1}{m-1} e^{-(m-1)A}(1-e^{-A})^{N-m}\log_2 (N/m)\right]^{-1}. \qquad (18)$$

Fig. 2 show a plot of c_t versus N for three values of E_t: $E_t = 4kT$ (Ising's value), $E_t = kT \ln 2$ (for which $p(n \geq n_t) = p(n < n_t) = 1/2$) and $E_t \to 0$. It can be shown that c_t is minimized for $E_t \to 0$, giving

$$\min c_t = \frac{kT}{(N-1)\log_2\left(\frac{N}{N-1}\right)} \qquad (19)$$

corresponding to an efficiency

$$\max \eta = (N-1)\ln\left(\frac{N}{N-1}\right). \qquad (20)$$

For large values of N the expression for c_t approaches the ideal value $kT \ln 2$, in agreement with (2), while η approaches unity. A proof (not shown here) can be given that these limits are also approached, though more slowly, for all finite values of E_t; this behavior may be seen indicated in Fig. 2.

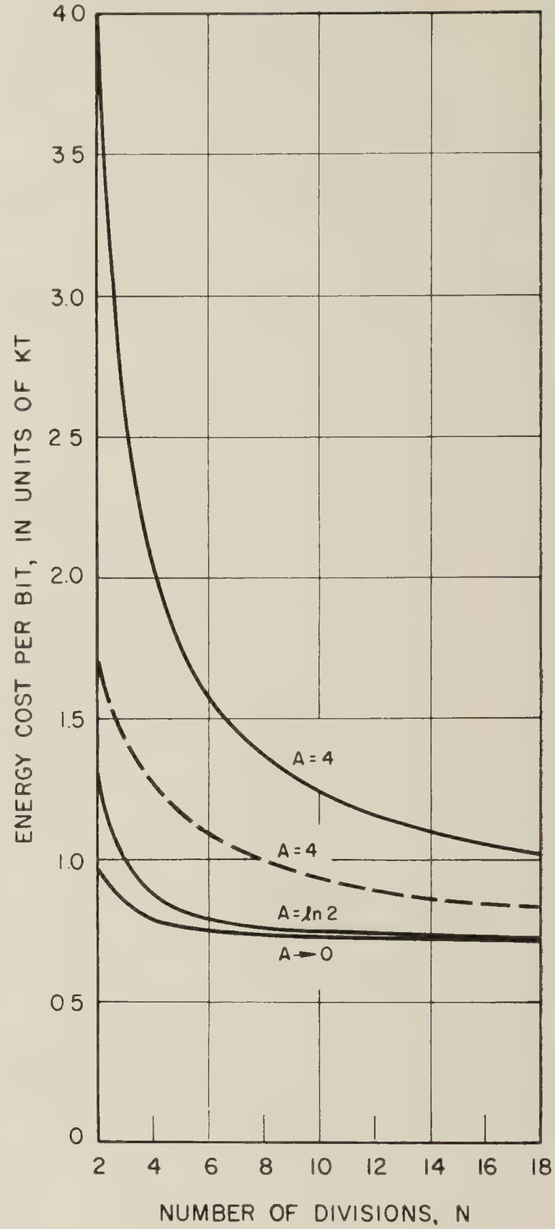


Fig. 2—Energy cost of basic time determination. Number of divisions (N) is a measure of accuracy of measurement. Signal or threshold energy (AkT) is a measure of reliability of measurement. Dotted curve is based on Shannon's channel capacity formula (for $A = 4$).

Length Measurement

We shall consider here the problem of determining the distance to some test object known to be in the field of observation. We first analyze an experimental procedure analogous to one described by Brillouin[10]. The total length L is considered to be divided into $N = L/\Delta L$ intervals, where ΔL is the desired accuracy of measurement. In order to discover now which one of the cells ΔL contains the particle, a small beam of light is directed at each one in turn. A cell is then presumed to be empty if a resonator placed behind it is excited by the light. If the resonator does not show an indication, however, it must be in the shadow of the particle which has absorbed the light. We again set a threshold level $E_t = n_t h\nu$ and consider that if an energy state $nh\nu$ is observed on a resonator, the resonator has received the light beam provided $n \geq n_t$ (cell empty), and has not if $n < n_t$ (cell contains the particle). In order that a resonator behind an empty

ell will always show an $n \geq n_t$, the light beam energy ised in scanning one cell is taken equal to E_t.

For the experiment to be successful the resonator behind he particle must not exceed E_t due to thermal fluctuations t the time of observation. Otherwise no information at ll is obtained since the cell containing the particle is ndistinguishable from the $N - 1$ remaining cells. The canning process may therefore have to be repeated once r several times before the particle position is determined. The probability of finding it on the first scan is, using (15),

$$p(n < n_t) = 1 - e^{-A}. \tag{21}$$

'or finding it on the second scan only, it is

$$p(n \geq n_t)p(n < n_t) = e^{-A}(1 - e^{-A}). \tag{22}$$

'inally, the probability of discovering the proper cell at he kth trial, after $k - 1$ unsuccessful scans, is

$$p(n \geq n_t)^{k-1}p(n < n_t) = e^{-(k-1)A}(1 - e^{-A}). \tag{23}$$

Now on each unsuccessful scan, Nn_t quanta are expended, vhile the average cost of the (final) successful test is

$$(1 + 2 + 3 + \cdots + N)n_t/N = \frac{N+1}{2} n_t \text{ quanta} \tag{24}$$

ince each cell position is equally likely. The average otal energy expenditure for a successful observation $n < n_t$) is therefore

$$\bar{E} = \sum_{k=1}^{\infty} e^{-(k-1)A}(1 - e^{-A})\left[(k-1)N + \frac{N+1}{2}\right]E_t$$
$$= [(N+1)/2 + N/(e^A - 1)]E_t. \tag{25}$$

Once $n < n_t$ is observed for a particular cell, that cell s known to contain the particle. The *a posteriori* unertainty, I_1, is therefore zero and hence the information ain is

$$\Delta I = \log_2 N. \tag{26}$$

The average cost is therefore

$$c_L = \frac{\bar{E}}{\Delta I} = \frac{(N+1)/2 + N/(e^A - 1)}{\log_2 N} AkT \text{ ergs per bit}. \tag{27}$$

t can be shown that c_L minimizes for $A \to 0$ giving

$$\min c_L = \frac{N}{\log_2 N} kT. \tag{28}$$

Plots of c_L, again for $E_t = 4kT$, $kT \ln 2$ and $E_t \to 0$, re shown in Fig. 3. Comparison with Fig. 2 shows that vhile c_L and c_t are roughly comparable in the region of ow accuracy (small N), the length measurement cost ecomes increasingly large as the accuracy is increased. The reason for the relative inefficiency of the length bservation is not difficult to see: the number of signals ent out and absorbed is of the order N, while in the case f the time measurement only a single signal pulse is nvolved. The high c_L values found are not believed to be asic to the problem of length measurement, however. A more efficient procedure, for example, would be to reduce the distance to a time measurement analogously to the operation of a pulse radar set: a pulse is sent from 0 to the particle position P and the time interval between transmission of the pulse and receipt of the reflected pulse is observed as described in the section on time measurement; the distance $\overline{0P}$ is then obtained from a knowledge of the speed of light.

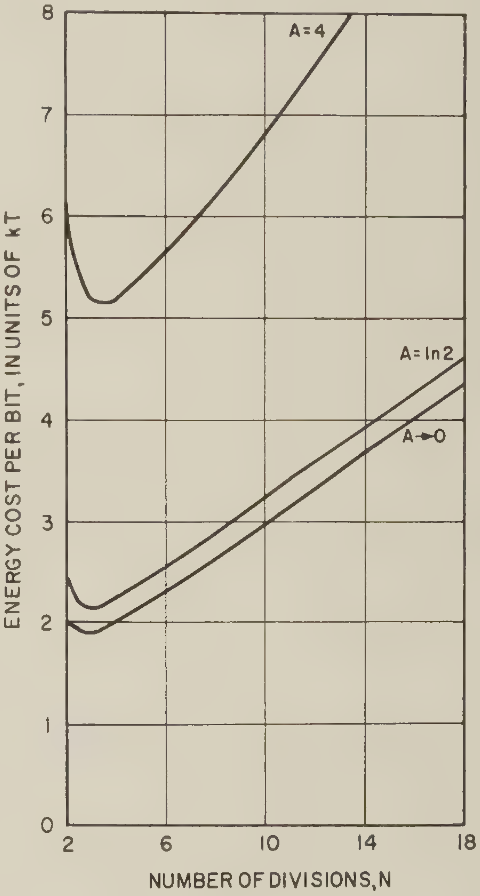


Fig. 3—Energy cost of a length determination.

Discussion

It is of interest to compare the results obtained in the section on time measurement with Shannon's classical channel formula. In slightly modified form, this relation states that the maximum information gain obtainable in a time t with a signal energy E and a channel bandwidth W is

$$\Delta I = Wt \log_2 \left(1 + \frac{E/t}{n_0 W}\right) \text{ bits} \tag{29}$$

where n_0 is the noise power spectral density which has the minimum value kT. The minimum energy cost is therefore

$$c_S = \frac{2E}{N \log_2 \left(1 + \frac{2E}{NkT}\right)} \text{ ergs per bit} \tag{30}$$

where $2\ Wt$, the number of independent samples which can be transmitted in time t with bandwidth W, has been

denoted by N, since this number is roughly analogous to the N (number of detectors or divisions) used in the two preceding sections. For $N \to \infty$, c_S is minimized to

$$\min c_S = kT \ln 2 \text{ ergs per bit} \tag{31}$$

a result which has been pointed out before[18]. A plot of c_S is shown in Fig. 2 for $E = 4kT$. Although the precise significance of this curve for the small N-values shown may be open to some discussion since (29) holds strictly only for a long signal sequence, it does bring out the higher efficiency obtainable with (ideal) coding than with a straightforward measurement procedure.[19]

In addition to the quantitative agreement between (31) and the result obtained in the section on time measurement, there is also a broader similarity between Shannon's information cost (30) and that derived here (18). In both cases, cost is minimized by reducing the reliability of a received signal indication to a minimum, by using a vanishingly small value of signal energy ($E \to 0$), and by making the number of independent samples or divisions very large ($2Wt \to \infty$, $N \to \infty$). As a practical matter, these low-cost conditions are usable for communication systems employing efficient wide-band coding and modulation schemes; for the case of physical measurements, however, they are not the ones commonly used in the laboratory since awkward repetition procedures would be required to obtain significant amounts of reliable information. There thus appears to be an important distinction between communication systems which allow ingenious coding of the information, and measurement procedures in which the quantity to be determined has t be measured in pretty much its nature-given form[19].

[18] See, for example, J. A. Felker, "A link between information and energy" Proc. IRE, vol. 40, p. 728; June, 1952. Felker, however, associates E with the battery or power supply energy required to amplify a signal. This is in disagreement with the more fundamental viewpoint adopted here, as well as in the paper by Brillouin, who states in connection with the timing scheme of the section on time measurement: "We may think of using a tuned receiver, but usually an amplifier is needed. This is just an auxiliary device to increase the power and we are not going to consider the energy required in the amplifying system itself" (*J. Appl. Phys.*, vol. 24, p. 1161; 1953. Cf. also F. P. Adler, "Comments on 'Figure of merit for communication devices'," Proc. IRE, vol. 42, p. 1191; July, 1954.

[19] For an interesting discussion of the differences between communication and measurement systems see P. H. Blundell, "The definition of rate of information in the presence of noise," p. 27 of "Communication Theory" (papers read at a Symposium on Communication Theory held at London), Butterworth's Scientific Publications (1953).

Appendix

Time Measurement Without Threshold Level

As was stated in a footnote to the text of the sectio on time measurement, it may seem reasonable that mor information could be extracted if, instead of merel observing whether or not the energy of a resonator ex ceeded some threshold level, the actual energy valu were to be noted. Suppose, then, that an energy leve $n_i h\nu$ is observed on the ith detector. The probability c observing this level due to noise alone (hypothesis H_0) i

$$p(n_i/H_0) = Ke^{-n_i h\nu/kT}.$$

If the detector has received the signal (hypothesis H_1) this probability becomes

$$p(n_i/H_1) = \begin{cases} Ke^{-(n_i - s)h\nu/kT} & \text{for} \quad n_i \geq s \\ 0 & \text{for} \quad n_i < s \end{cases}$$

where $sh\nu$ is energy of signal pulse. Using Bayes' theorem and assuming *a priori* probabilities $p(H_0) = p(H_1) = 1/2$ the following *a posteriori* probabilities are obtained:

$$p(H_0/n_i) = \begin{cases} \dfrac{1}{1 + e^{sh\nu/kT}} & \text{if} \quad n_i \geq s \\ 1 & \text{if} \quad n_i < s \end{cases}$$

$$p(H_1/n_i) = \begin{cases} \dfrac{1}{1 + e^{-sh\nu/kT}} & \text{if} \quad n_i \geq s \\ 0 & \text{if} \quad n_i < s. \end{cases}$$

The observation of all Nn_i's will therefore divide th resonators into two groups: those which are known not t have received the signal ($n_i < s$) and those which ma have received it ($n_i \geq s$). Since $p(H_1/n_i)$ for $n_i \geq s$ doe not depend on n_i, the detectors in the second group ar indistinguishable from each other; an observed energ level $(s + 10^6)\ h\nu$ is no more indicative of a receive signal than a level of $sh\nu$. Signal energy value acts effec tively as a threshold or dividing level and rest of analysi proceeds as in the section on time measurement.

Some Remarks on Statistical Detection*

W. L. ROOT† AND T. S. PITCHER†

Summary—A particular type of communications detection problem is considered: the problem of specifying a detector to decide which one of two sure signals is being transmitted when the signals are perturbed randomly both by Gaussian noise and multipath transmission. If the delays in the various channels are known, but the strengths are random, a maximum likelihood detector may be specified by methods which are a simple extension of known methods. If the delays are random, the problem is more difficult. One possible solution is first to estimate certain channel parameters from the received signal and then to use these estimates in a likelihood test. It is shown how to make consistent unbiased estimates for appropriate channel parameters under certain assumptions on the nature of the signal.

Introduction

In the last few years the general problem of detecting signals in the presence of noise has been recognized in its idealized form to be statistical, and a good deal of theoretical work has been done on it (see *e.g.* [4] and [5]. One common type of detection problem is to establish a criterion for determining whether or not a known signal is present in noise, another is to establish a criterion for determining which one of a collection of known possible signals is present in noise. Usually the noise is represented mathematically as a stationary Gaussian stochastic process independent of the signal which is simply added to the signal by the channel. In this paper we look at problems of the latter type but we consider cases where the signal is randomly perturbed in other ways than just by the addition of Gaussian noise. The particular communications problem which motivated this work is that of receiving signals (which are sent in a two-letter alphabet) over a radio channel in which the signals are distorted both by additive noise and by multipath propagation with in general random (or at least unknown) characteristics. Before formulating any examples, we discuss briefly and very generally the point of view which we adopt in treating them.

Let $X_0(t)$, $X_1(t)$ be real-valued functions of a real variable t, $0 \leqq t \leqq T$. These functions represent two possible known signals one of which is to be transmitted during a time interval of duration T. Let P represent a channel randomly perturbing the transmitted signal. We ask for a procedure which will, from a knowledge of

$$s(t) = P(X_i(t)), \qquad i = 0, 1,$$

lead to a reasonable judgment as to whether i is equal to 0 or 1; *i.e.*, we ask for a detector which will tell as well as possible from the received signal which of the two possible transmitted signals was actually sent.

In statistical language, $s(t)$ is one of two stochastic processes, depending upon whether $i = 0$ or 1, and we want a statistical test which will allow us from observing a realization of $s(t)$ to decide which process it belongs to.

This problem is one of testing statistical hypotheses and making a decision. We do not here want to enter into the difficult matter of optimum statistical decisions, but rather want to discuss tests for certain specific cases with the understanding that the method of making the decision is fixed. Our detection procedure can be described in general as follows: Let $(z_1, z_2, \cdots)$ be a finite or countably infinite set of measurable functions on the underlying probability space, *i.e.*, under hypothesis i, $i = 0$ or 1, $(z_1, z_2, \cdots)$ is a vector-valued random variable. Let probability densities $f_0(z_1, \cdots, z_n)$, $f_1(z_1, \cdots, z_n)$ be defined for both hypotheses for all n if there are infinitely many z's and for n equal to the number of z's if there are finitely many. Form the likelihood ratio

$$l_n = \frac{f_0(z_1, \cdots, z_n)}{f_1(z_1, \cdots, z_n)},$$

and suppose it converges to l if $n \to \infty$. Then a realization $s(t)$ of $P(X_i(t))$ will assign values to the z's and hence to l; choose $i = 0$ or 1 according as l is $>$ or $<$ 1. Following Grenander's [1] terminology we call the z's "observable coordinates".

In the first section we consider detection from the point of view just outlined for the cases in which the received signal $s(t)$ is given by

1) $\quad s(t) = x_i(t) + n(t), \qquad i = 0, 1,$

2) $\quad s(t) = x_i(t) + \alpha x_i(t - \tau_0) + n(t), \qquad i = 0, 1,$

where $n(t)$ is a stationary Gaussian process with mean zero and α is a random variable independent of $n(t)$. The first case is known but since it applies to what follows we have summarized the pertinent results. We have followed Grenander [1] in doing this. We have also shown that a particularly simple type of singular case of (1) does not arise if $n(t)$ is filtered white noise. The second case can be easily generalized to one with more delay terms.

In the second section we again consider case (2) and then the much more general case,

$$s(t) = \int x_i(t - \tau)\, dF(\tau) + n(t)$$

where $dF(\tau)$ is a Stieltjes measure. In this section we regard the channel parameters (α and dF) as having

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

† Staff Members, M.I.T., Lincoln Laboratory, Lexington, Mass.

unknown probabilistic behavior and the technique is to find consistent linear estimates for them from $s(t)$ and use these estimates to simplify the statistical hypotheses test problem.

I. Maximum Likelihood Test for Signals with Known Delay

In this section we first review briefly the problem of choosing i when $s(t)$ has the representation

$$s(t) = x_i(t) + n(t), \qquad i = 0, 1, \qquad 0 \leq t \leq T$$

$n(t)$ being stationary Gaussian noise with mean zero and autocorrelation $r(s, t) = \rho(s - t)$. Let $\{\varphi_k(t)\}$ be a set of orthonormal eigenfunctions and $\{\lambda_k\}$ the corresponding eigenvalues of

$$\lambda \int_0^T r(s, t)\varphi(t)\, dt = \varphi(s).$$

Then by a theorem of Karhunen and Loeve [6, pp. 327, 328]

$$n(t) = \text{l.i.m.} \sum_k z_k\varphi_k(t)$$

where

$$z_k = \int_0^T n(t)\varphi_k(t)\, dt.$$

The z_k's are random variables which satisfy

$$E\{z_k\bar{z}_m\} = \delta_{km}\frac{1}{\lambda_k}$$

$$E\{z_k\} = 0.$$

where the bar denotes complex conjugate, $E\{\cdot\}$ denotes mathematical expectation or average and δ_{km} is equal to 1 if $k = m$ and zero otherwise. Now since $n(t)$ is a Gaussian random process the z_k have a joint Gaussian distribution and hence are independent. We introduce the notation $a_k = (\varphi_k, x_0)$ and $b_k = (\varphi_k, x_1)$, where

$$(\varphi_k, x_0) \text{ is the "inner product" } \int_0^T \varphi_k(t)\overline{x_0(t)}\, dt,$$

and use as observable coordinates

$$y_k = (\varphi_k, s) = a_k + z_k \quad \text{if} \quad i = 0$$
$$= b_k + z_k \quad \text{if} \quad i = 1, \quad k = 1, 2, \cdots, \quad \text{and}$$

$\hat{y}_k = (\psi_k, s)$, where the φ_k's and ψ_k's form a complete orthonormal set.

If there is a function $\psi(t)$ orthogonal to $\{\varphi_k(t)\}$, *i.e.* $(\psi, \varphi_k) = 0$, $k = 1, 2, \cdots$, for which $(x_0, \psi) \neq (x_1, \psi)$, the problem is said to be *extreme singular*. Then since $\psi(t)$ is orthogonal to $n(t)$ with probability one, the statistic (s, ψ) is equal with probability one to (x_0, ψ) if $i = 0$ and to (x_1, ψ) if $i = 1$, thus providing perfect detection. In practice however $\{\varphi_k(t)\}$ usually spans the whole space so the extreme singular case cannot happen. This is so in particular if $n(t)$ is filtered white noise (see Appendix

We consider now the remaining case, in which t difference of the mean-value functions, $x_1(t) - x_0(t)$, li in the Hilbert space spanned by the eigenfunctio $\{\varphi_k(t)\}$. If ψ is orthogonal to $\{\varphi_k\}$ then (s, ψ) is a consta independent of i, so we need only consider the observab coordinates (s, φ_k), $k = 1, 2, \ldots$.

Now

$$E\{y_k\} = a_k \quad \text{if} \quad i = 0$$
$$= b_k \quad \text{if} \quad i = 1$$

$$\text{var}\,\{y_k\} = \text{var}\,\{a_k\} = \frac{1}{\lambda_k} \quad \text{for} \quad i = 0 \quad \text{or} \quad i = 1.$$

where var $\{y_k\}$ means the variance of the random variab y_k.

Hence

$$f_0(y_1, \cdots, y_n) = \frac{\sqrt{\lambda_1 \cdots \lambda_n}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_1^n \lambda_k(y_k - a_k)\right.$$

and $f_1(y_1, \cdots, y_n)$ is the same except b_k replaces a_k. Th

$$l_n(\omega) = \frac{f_1(y_1, \cdots, y_n)}{f_0(y_1, \cdots, y_n)}$$
$$= \exp\left\{-\frac{1}{2}\sum_1^n \lambda_k(b_k^2 - a_k^2) - \sum_1^n y_k(a_k - b_k)\lambda_k\right\}.$$

Let

$$f_n(t) = \sum_1^n (a_k - b_k)\lambda_k\varphi_k(t),$$

then

$$\log l_n(\omega) = \frac{1}{2}\sum_1^n \lambda_k(a_k - b_k)(a_k + b_k) - \sum_1^n \lambda_k(a_k - b_k)$$
$$= \frac{1}{2}\sum_1^n \lambda_k(a_k - b_k)\int_0^T \varphi_k(t)(x_0 + x_1)\, dt$$
$$- \sum_1^n \lambda_k(a_k - b_k)\int_0^T \varphi_k(t)s(t)$$
$$= \int_0^T f_n(t)\left(\frac{x_0 + x_1}{2} - s(t)\right) dt.$$

If

$$\sum_1^\infty \lambda_k(a_k - b_k)^2 < \infty$$

it can be shown [1, pp. 215–216] that log $l_n(\omega)$ converg to a finite value with probability one under eith hypothesis. If

$$\sum_1^\infty \lambda_k(a_k - b_k)^2 = +\infty$$

then log $l_n(\omega)$ converges to $+\infty$ in probability wi respect to hypothesis 1 and $-\infty$ in probability wi respect to hypothesis 0; *i.e.*, the singular case obtains.

Thus a maximum likelihood test consists in determining whether

$$\lim_{n\to\infty} \int_0^T f_n(t)\left(\frac{x_0(t) + x_1(t)}{2} - s(t)\right) dt$$

is less than or greater than zero; if the limit is $+\infty$ or $-\infty$ the test gives the correct hypothesis with probability one. Now,

$$\int_0^T r(s - t)f_n(t)\,dt = \int_0^T \sum^{\infty} \frac{\varphi_k(s)\varphi_k(t)}{\lambda_k} f_n(t)\,dt$$

$$= \sum_1^n (a_k - b_k)\varphi_k(s)$$

by Mercer's theorem. Hence, since $x_0 - x_1$ is of integrable square on $(0, T)$,

$$\underset{n}{\text{l.i.m.}} \int_0^T r(s - t)f_n(t)\,dt = x_0(s) - x_1(s).$$

If now in addition f_n converges in mean square to a function f, then

$$\int_0^T r(s - t)f(t)\,dt = x_0(s) - x_1(s)$$

for a.e.s. In this case

$$\lim_{n\to\infty} \int_0^T f_n(t)\left[\frac{x_0(t) + x_1(t)}{2} - s(t)\right] dt$$

$$= \int_0^T f(t)\left[\frac{x_0(t) + x_1(t)}{2} - s(t)\right] dt$$

with probability one; thus the test function f must be a solution of the integral equation above. Conversely if the integral equation has a solution f of integrable square, then this f will serve as a test function [1, pp. 218–219]. Let

$$Rf(s) = \int_0^T r(s - t)f(t)\,dt.$$

Now if the spectrum of r is broad and flat, $f = x_0 - x_1$ is nearly a solution. That is, under this circumstance the maximum likelihood test is nearly a correlation test.

If delay terms with fixed delay are present as well as the additive noise, it is still possible to obtain the likelihood ratio using the same observable coordinates as above. We show this only for the simplest such case. Let

$$s(t) = x_i(t) + \alpha x_i(t - \tau) + n(t)$$

where τ is known. If $i = 0$ the joint density of $y_1, \cdots, y_N$ is

$$\int_{-\infty}^{\infty} f_\alpha(\beta) \prod_{k=1}^{N} f_k(y_k - a_k - \beta c_k)\,d\beta$$

where

$$c_k = \int_0^T \varphi_k(t)x_0(t - \tau)\,dt,$$

where f_α is the density function of α and f_k is the Gaussian density function with mean 0 and variance $1/\lambda_k$. The density for $i = 1$ is obtained by substituting b_k and for a_k and

$$d_k = \int_0^T \varphi_k(t)x_1(t - \tau)\,dt$$

for c_k. By a Martingale theorem the ratio of the two converges as $N \to \infty$ so in theory one can use them to perform a maximum likelihood test. [2, p. 348ff].

II. Estimation of Channel Parameters

In many situations it is unreasonable to assign probability distributions to the channel parameters. Hence there is some interest in developing a procedure to estimate these parameters and then use these estimates to reduce the problem to a simpler one of hypotheses testing, that of choosing the mean of a Gaussian process.

We first treat the case

$$s(t) = x_i(t) + \alpha x_i(t - \tau) + n(t)$$

with τ fixed by finding a minimum variance unbiased estimate for α of the form $\alpha^* = c + (s, f)$. For α^* to be unbiased we must have

$$E_\alpha(\alpha^*) = k + (x_i, f) + \alpha(x_{i,\tau}, f) = \alpha$$

for all α, where E_α indicates the expectation if α is true and we have written $x_{i,\tau}$ for x_i delayed by τ. The above is true if and only if

$$(x_0, f) = (x_1, f) = -k$$

and

$$(x_{0,\tau}, f) = (x_{1,\tau}, f) = 1.$$

Also

$$\operatorname{var}(\alpha^* - \alpha) = \operatorname{var}(f, n)$$

$$= \int_0^T \int_0^T r(s - t)f(s)f(t)\,ds\,dt = (Rf, f),$$

where

$$Rf(s) = \int_0^T R(s - t)f(t)\,dt.$$

If for some a, b, and c, f satisfies the constraints above and

$$Rf = ax_{0,\tau} + bx_{1,\tau} + c(x_0 - x_1)$$

then

$$\alpha^* = (s, f) - (x_0, f)$$

is a minimum variance estimate. For if

$$\hat{\alpha}^* = (h, f) - (h, x_0)$$

is another unbiased estimate; we have

$$\operatorname{var}(\hat{\alpha}^* - \alpha) = \operatorname{var}(h, n) = (Rh, h) = (Rf, f)$$

$$+ 2(Rf, h - f) + (R(h - f), h - f)$$

but the last term is nonnegative because of the nature of the operator R and the middle term is zero since

$$(Rf, h - f) = a(x_{0,\tau}, h - f) + b(x_{1,\tau}, h - f) + c((x_0 - x_1), h - f) = 0.$$

In particular if f_1, f_2, and f_3 are solutions of $Rf_1 = x_{0,\tau}$, $Rf_2 = x_{1,\tau}$, $Rf_3 = x_0 - x_1$ then $af_1 + bf_2 + cf_3$ will give a minimum variance estimate if a, b, and c satisfy the albegraic equations

$$(af_1 + bf_2 + cf_3, x_0 - x_1) = 0$$

$$(af_1 + bf_2 + cf_3, x_{0,\tau}) = 1$$

$$(af_1 + bf_2 + cf_3, x_{1,\tau}) = 1.$$

The above model is a special case of a more general one which we consider in the following paragraphs. There we show a method which is easier to apply than the above for obtaining unbiased consistent estimates of channel parameters. These estimates are not in general of minimum variance, however.

Let

$$s(t) = \int x_i(t - \tau)\, dF(\tau) + n(t)$$

dF being any Stieltjes measure confined to a reasonable interval. We assume that

$$\rho_{00}(t) = \rho_{11}(t) = \sum_{i=0}^{N} a_i t^i$$

$$\rho_{10}(t) = \rho_{01}(t) = \sum_{i=0}^{N} b_i t^i$$

and write $c_i = a_i + b_i$. We assume below that

$$\int_0^T x_i(t - \sigma)x_j(t - \tau)\, dt = \int_0^T x_i(t)x_j(t - (\tau - \sigma))\, dt = \rho_{ij}(\tau - \sigma).$$

There is a class of functions $\{x(t)\}$ for which this is true, of course, and it is true asymptotically for any function which might represent a conventional signal. Then,

$$(s, x_{0,\tau} + x_{1,\tau}) = \sum_{i=0}^{N} c_i \int (\tau - \sigma)^i\, dF(\sigma) + (n, x_{0,\tau} + x_{1,\tau}).$$

Forming the inner product above for $N + 1$ different values of τ and expanding the terms $(\tau - \sigma)^i$ gives a set of equations of the form

$$s_j = (s, x_{0,\tau_j} + x_{1,\tau_j}) = \sum_{k=0}^{N} a_{jk} \int \tau^k\, dF(\tau) + (n, x_{0,\tau_j} + x_{1,\tau_j})$$

where each a_{jk} is a linear combination of the c's. Setting the noise term to zero and solving the resultant set of equations gives unbiased estimates $\mu_0, \cdots, \mu_N$ for the first N moments of dF and leads to the approximations

$$(s, x_0) = \sum_{k=0}^{N} a_k\mu_k + (s, n) \quad \text{if} \quad i = 0$$

$$= \sum_{k=0}^{N} b_k\mu_k + (s, n) \quad \text{if} \quad i = 1.$$

Thus we are left with the problem of testing for the me of a Gaussian random variable. In Appendix II we pro the consistency of these estimates as $T \to \infty$. If T remai fixed and successive estimates are averaged, the avera estimates are also consistent.

We illustrate the above technique in a simple cas Suppose

$$\rho_{00}(t) = a_0 + a_1 t, \qquad \rho_{01}(t) = b_0 + b_1 t,$$

then

$$\mu_0 = \frac{(s, x_0 + x_1) - (s, x_{0,\sigma} + x_{1,\sigma})}{\sigma c_1}$$

$$\mu_1 = \frac{(\sigma c_1 - c_0)(s, x_0 + x_1) - c_0(s, x_{0,\sigma} + x_{1,\sigma})}{\sigma^2 c_1}.$$

These estimated moments of dF may then be used, f example, to give estimated values of the mean of t normally-distributed statistic (s, x_0). In particular:

Estimated mean of $(s, x_0) = a_0\mu_0 + a_1\mu_1$ if x_0 was sent,

$= b_0\mu_0 + b_1\mu_1$ is x_1 was sent.

The hypothesis test remaining is to choose which of the mean values is most likely.

In the problem considered above where

$$s(t) = x_i(t) + \alpha x_i(t - \tau_0) + n(t)$$

we have $dF(0) = 1$, $dF(\tau_0) = \alpha$. Then by equating t estimated moments to the true moments we have

$$\mu_0 = \int dF(\tau) = 1 + \alpha$$

$$\mu_1 = \int dF(\tau) = \alpha\tau_0$$

whence $1 - \mu_0$ is an unbiased estimate for α. Then usi this estimate for α in the expression for $s(t)$ we have le a hypothesis testing problem which approximates t original one but is simpler; it is in fact precisely t problem discussed in Section I.

Appendix I

Suppose $n(t)$ is filtered white noise; *i.e.*,

$$n(t) = \int_{-\infty}^{\infty} C(u)\, d\xi(t + u),$$

where $C(u)$ is the filter impulse response which vanish for $u > 0$ and which we assume is integrable and squa integrable, ξ is a stochastic process with orthogor increments and the integral involved is a stochas integral [2, p. 539]. Then

$$\rho(t-s) = E(n(t)\overline{n(s)}) = \int_{-\infty}^{\infty} C(u-t)\overline{C(u-s)}\, du.$$

Theorem: The eigenfunctions of

$$\int_0^T \rho(s-t)f(t)\, dt = \frac{1}{\lambda} f(s)$$

pan the whole L_2 space.
Proof: Let

$$Rf(s) = \int_0^T \rho(s-t)f(t)\, dt$$

nd suppose $Rf = 0$ for some f so that $(Rf, f) = 0$, then

$$\begin{aligned} 0 &= \int_0^T dt\, f(t) \int_0^T ds\, \rho(t-s)\overline{f(s)} \\ &= \int_0^T dt \int_0^T ds \int_{-\infty}^{\infty} du\, C(u-t)\overline{C(u-s)} f(t)\overline{f(s)} \\ &= \int_{-\infty}^{\infty} du \left(\int_0^T dt\, C(u-t)f(t) \right) \overline{\left(\int_0^T ds\, C(u-s)f(s) \right)} \\ &= \int_{-\infty}^{\infty} du \mid S\hat{f} \mid^2 (u) \end{aligned}$$

here

$$S\hat{f}(u) = \int_{-\infty}^{\infty} C(u-s)\hat{f}(s)\, ds$$

and $\hat{f}$ is f on $(0, T)$ and 0 outside.

But the Fourier transform of Sf is the product of the Fourier transforms of C and f and the transform of C cannot vanish on any set of positive measure by a theorem f Paley and Wiener [3, p. 739] so the transform of f anishes everywhere; *i.e.*, $f = 0$. Now suppose g is rthogonal to all the eigenfunctions and hence to all unctions of the form Rh. Then $(Rg,h) = (g,Rh) = 0$ for ll h so $Rg = 0$ so $g = 0$.

Appendix II

In order to investigate the consistency of the test for the noments of dF we suppose that for all τ in the interval here dF has positive measure and for $\tau = 0$ we have

$$\rho_0(\tau) = \lim_T \frac{1}{T} \int_0^T x_0(t+\tau)x_0(t)\, dt = \lim_T \frac{1}{T} \int_0^T x_1(t+\tau)x_1(t)\, dt = \sum_{i=0}^{N} a_i \tau^i$$

nd

$$\rho_1(\tau) = \lim_T \frac{1}{T} \int_0^T x_0(t+\tau)x_1(t)\, dt = \lim_T \frac{1}{T} \int_0^T x_1(t+\tau)x_0(t)\, dt = \sum_{i=0}^{N} b_i \tau^i$$

and for each T we choose polynomials

$$\rho_0^T(\tau) = \sum_{i=0}^{N} a_i(T)\tau^i$$

and

$$\rho_1^T(\tau) = \sum_{i=0}^{N} b_i(T)\tau^i$$

in such a way that $a_i(T) \to a_i$ and $b_i(T) \to b_i$. We choose $\tau_0, \cdots, \tau_n$ so that the equations,

$$\lim \frac{1}{T} \int_0^T x_0(t)(x_0(t+\tau_j) + x_1(t+\tau_j))\, dt = \sum_{k=0}^{N} a_{jk} \int \tau^k dF(\tau)$$

will have a solution; that is, matrix $A = (a_{ij})$ will have an inverse. Let $s(T)$ be vector whose jth component is

$$\frac{1}{T} \int_0^T s(t)[x_0(t+\tau_j) + x_1(t+\tau_j)]\, dt$$

and let $A(T)$ be the matrix made up in the same way that A is using $a_i(T)$ and $b_i(T)$ instead of a_i and b_i. Then for large enough T, $A(T)^{-1}$ exists since $A(T) \to A$ and we set $\mu(T) = A(T)^{-1} s(T)$.
Theorem: If $r(\infty) = 0$ then the estimates

$$m_0^*(T) = \sum_{i=0}^{N} a_i(T)\mu_i(T) \quad \text{and} \quad m_1^*(T) = \sum_{i=0}^{N} b_i(T)\mu_i(T)$$

for the means of (s, x_0) and (s, x_1), under the assumption that the autocorrelations over the interval 0 to T are given by ρ_0^T and ρ_1^T, are consistent.
Proof: Let $n(T)$ be the vector whose jth component is

$$\frac{1}{T} \int_0^T n(t)[x_0(t+\tau_j) + x_1(t+\tau_j)]\, dt$$

and let $\bar{\mu}(T) = A(T)^{-1} (s(T) - n(T))$, then

$$\begin{aligned} &\mathrm{var}\,(m_0^*(T) - m_0(T)) \\ &= \mathrm{var}\left(\sum a_i(T)\mu_i(T) - \int \rho_0^T(\tau)\, dF(\tau) \right) \\ &= \mathrm{var}\Big[\sum a_i(T)(\mu_i(T) - \bar{\mu}(T)) \\ &\quad + \sum a_i(T)\left(\bar{\mu}_i(T) - \int \tau^i\, dF(\tau)\right) \\ &\quad + \int (\sum a_i(T)\tau^i - \rho_0^T(\tau))\, dF(\tau) \Big]. \end{aligned}$$

For large enough T

$$\begin{aligned} &\mathrm{var}\,(\sum a_i(T)(\mu_i(T) - \bar{\mu}_i(T))) \\ &\le 2\, \mathrm{m_iax}_i \mid a_i \mid^2 N^2\, \mathrm{m_jax}_j\, \mathrm{var}\left(\frac{1}{T} \int_0^T n(t)[x_0(t+\tau_j) + x_1(t+\tau_j)]\, dt \right) \end{aligned}$$

so it will be sufficient for this term to show that

$$\frac{1}{T^2}\int_0^T\int_0^T r(s-t)x_a(t-\tau)x_b(s+\tau)\,ds\,dt \to 0$$

for any a, b, and τ. By Schwarz's inequality the square of this is dominated by

$$\frac{1}{T^2}\left(\int_0^T\int_0^T r^2(s-t)\,ds\,dt\right)$$

$$\cdot\left(\frac{1}{T}\int_0^T x_a^2(t+\tau)\,dt\right)\left(\frac{1}{T}\int_0^T x_b^2(s+\tau)\,ds\right)$$

$$= 0\left(\frac{1}{T^2}\int_0^T\int_0^T r^2(s-t)\,ds\,dt\right)$$

since the last two terms converge to $\rho_0(0)$. Finally

$$\frac{1}{T^2}\int_0^T\int_0^T r^2(s-t)\,ds\,dt = \int_0^1\int_0^1 r^2(T(u-v))\,du\,dv$$

which goes to zero by the dominated convergence theorem. Since $A(T)^{-1} \to A^{-1}$,

$$\bar{\mu}_i(T) \to \int \tau^i\,dF(\tau)$$

so the second term goes to zero, and the third term goes to zero by the dominated convergence theorem whi completes the proof.

Suppose that on each interval MT to $MT + T$ one the signals x_0 or x_1 is sent and we are required to deci at each MT which has been sent in the previous interv If ρ and r are as above and for each M we choose $a_i(M$ and $b_i(MT)$ to make the polynomials $\sum a_i(MT)\tau^i$ ar $\sum b_i(MT)\tau^i$ fit ρ_0^{MT} and ρ_1^{MT} at chosen points $\sigma_0, \cdots,$ so that $a_i(MT) \to a_i$ and $b_i(MT) \to b_i$ then the estimat

$$m_0^* = \sum a_i(MT)\mu_i(MT)$$

$$m_1^* = \sum b_i(MT)\mu_i(MT)$$

are easily seen to be consistent.

Bibliography

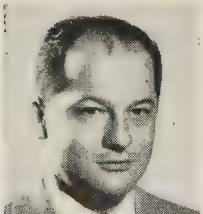
[1] Grenander, U., "Stochastic process and statical inference," *Ark for Matematik Band*, B. 1, H. 3. (1950), p. 195.
[2] Doob, J. L., *Stochastic Processes.* New York, John Wiley ar Sons, 1953.
[3] Valley, G. E., and Wallman, H., *Vacuum Tube Amplifiers.* Ne York, McGraw-Hill Book Company, 1948.
[4] Davis, R. C., "On the detection of sore signals in noise," *J. App Phys.*, Vol. 25 (1954), p. 76.
[5] Reich, E., and Swerling, P., "The detection of a sine wave Gaussian noise," *J. Appl. Phys.*, Vol. 24 (1953), p. 289.
[6] Levy, P., *Processus Stochastiques.* Note by M. Loeve. Par Gauthier-Villars, 1948.
[7] Price, R., "The detection of signals perturbed by scatter ar noise," *Transactions of the IRE*, Vol. PGIT-4 (1954), p. 163.

Correction

F. L. Stumper's paper, "A Bibliography of Information Theory (Communication Theory—Cybernetics)," which appeared in TRANSACTIONS OF THE IRE, Vol. IT-1, No. 2, pp. 31–47; September, 1955, is a supplement to a paper of the same title which appeared in a previous issue of TRANSACTIONS, Vol. IT-2; November, 1953.

Contributors

F. P. Adler was born in Vienna, Austria on March 29, 1925. He received a B.S. degree in electrical engineering from the University of California in February, 1945. In 1948 he received the M.S. degree and in 1950 a Ph.D. degree, both from the California Institute of Technology.

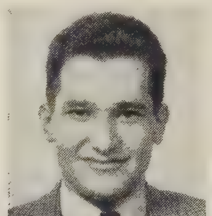


F. P. ADLER

From 1945 to 1947, he worked in the Consulting and Research Laboratories of the General Electric Co., on problems in electronic control, sonar and gaseous conductors. During that time he also took the General Electric Advanced Engineering Course. For the past five years he has been with the Research and Development Laboratories of the Hughes Aircraft Co., working mostly on missile guidance. His position is that of senior staff physicist in the Systems Analysis Department.

Dr. Adler is a member of Tau Beta Pi and Sigma Xi.

J. J. Bussgang (S'52–M'55) was born in Lwow, Poland, on March 26, 1925. His education was interrupted by military service in the Polish forces with the British Army in Italy. After the war, he studied at the Politecnico di Torino, Italy, and at the University of London, England, fro which he received the B.Sc. degree in eng neering in 1949. He received the S.M. degr in electrical enginee ing from M.I.T. 1951 and the Ph. degree in appli physics from Ha vard in 1955.



J. J. BUSSGANG

In 1950 and 195 he was a researc assistant at th M.I.T. Research La oratory of Electro ics. From 1952 1955, he was asso ated with the M.I. Lincoln Laboratory, first as a staff memb and later as a consultant. In 1954, he was the research staff of the Gordon McK

aboratory of applied science at Harvard. t present he is a senior member of the echnical staff of the new RCA Aviation ystems Laboratory in Waltham, Mass. His major work has been in the fields of tatistical communications and radar.

Dr. Bussgang is a member of Sigma Xi, he American Physical Society, the British nstitution of Electrical Engineers and the ociety for Industrial and Applied Mathmatics.

M. Horowitz was born in Sumter, S. C., on September 16, 1920. During World War II he served as an officer in the sound and

M. Horowitz

flash observation units of the Field Artillery. He received the B.S. degree in 1947 from the College of the City of New York, and in 1950, he received the M.S. in physics from Ohio State University, Columbus.

He worked in the Low Temperature Laboratory in Columbus, Ohio, on measurement of electronic specific heats of metals, and obtained the Ph.D. in 1953.

Since that time, Dr. Horowitz has been a member of the staff of the Aerophysics Department of the Goodyear Aircraft Corp. in Akron, Ohio. His major research studies have been concerned with work on applications of analog computers. At the present time, Dr. Horowitz is engaged in radar information studies at Goodyear Aircraft.

Dr. Horowitz is a member of Phi Beta Kappa, the American Physical Society, and Sigma Xi.

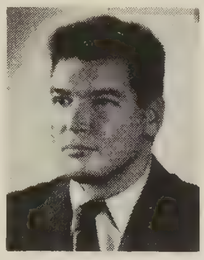
A. Johnson was born at Youngstown, Ohio, on March 15, 1924. He was a student at Kenyon College from 1943 to 1946, and then attended the College of Physicians and Surgeons in New York from 1946 to 1948. In 1948, he returned to Kenyon College and received the A.B. degree one year later. He received the M.A. degree in mathematics from Ohio State University in 1951. Since

A. Johnson

that time he has been employed at Goodyear Aircraft.

At present, Mr. Johnson is working toward a Ph.D. in mathematics at Case Institute of Technology. He is a member of Phi Beta Kappa and the American Mathematical Society.

D. Middleton (S'42–A'44–M'45–SM–53) was born on April 19, 1920, in New York, N.Y. He received the A.B. degree in 1942,



D. Middleton

the M.A. degree in 1945, and the Ph.D. degree in physics in 1947, all from Harvard University.

During World War II, he was a research associate at the Radio Research Laboratory at Harvard. and worked in the field of electronic counter-measures. From 1947 to 1949, he was a research fellow in electronics at the Electronics Research Laboratory of the division of applied science at Harvard. In 1949, he became assistant professor of applied physics in the same division. Since 1954, he has engaged in private consulting practice with industry and the armed services.

At present, Dr. Middleton's principal field of research is in statistical communication theory, including applications in electronics, electron physics, information theory, and system design and evaluation. He is also engaged in the study of various problems in applied mathematics which are related to these fields.

He is a fellow of the American Physical Society, a member of the American Mathematical Society, and the Society for Industrial and Applied Mathematics, Phi Beta Kappa, and Sigma Xi. He was a National Research Council predoctoral fellow in physics from 1946 to 1947.

T. S. Pitcher was born on April 17, 1926, in Wenatchee, Wash. After serving with the armed forces from 1944 to 1945, he at-

T. S. Pitcher

tended the University of Washington where he received the B.A. degree in 1949.

In 1953, he received the Ph.D. degree in mathematics from Massachusetts Institute of Technology. At that time, Dr. Pitcher joined the staff of Lincoln Laboratory of M.I.T.

W. L. Root was born on October 6, 1919, at Des Moines, Iowa. He received the S.B. in electrical engineering from Iowa State

W. L. Root

College in 1940, the S.M. in electrical engineering from M.I.T. in 1943 and the Ph.D. in mathematics from M.I.T. in 1952.

From 1940 to 1952 he was employed as a teaching assistant and instructor by M.I.T., except for a period of military service from 1943 to 1946. Since 1952, he has been a staff member of the Lincoln Laboratory of M.I.T.

R. A. Silverman (M'54) was born on June 29, 1926, in Boston, Mass. He received the A.B. from Harvard University

R. A. Silverman

in 1946, the M.A. from Columbia University in 1948, and the Ph.D. from Harvard in 1951.

For three years he was associated with the Lincoln Laboratory at M.I.T., and was also a research associate in the department of electrical engineering. Dr. Silverman is currently a research associate at the New York University Institute of Mathematical Sciences in the division of electromagnetic research.

IRE Professional Group on Information Theory Membership Directory

At the request of many members, and to further encourage the salutary associations which may result from a knowledge of whom to contact in specific regions, we are pleased to include herewith a Directory of PGIT members, according to Region and geographic subdivision. Business affiliations and addresses will be found in the 1955 IRE Directory. Further information may be obtained from your local chapter, your Professional Group Administrative Committee, or IRE Headquarters.

Laurin G. Fischer
Publications Chairman

Region I

Binghamton Section

Dickinson, M. M., Jr.
Hamburgen, Arthur
Hill, Y. M.
Marcy, H. T.
O'Connor, D. G.
Reeve, E. A.
Reynolds, A. C., Jr.

Boston Section

Baer, E. W., III
Baker, Adolph
Barry, J. G.
Bartelink, E. H. B.
Belkin, Arnold
Bishop, W. B.
Blanchard, R. L.
Bliss, Z. R.
Booton, R. C.
Briana, A. M.
Bronwell, A. B.
Brown, C. S.
Bussgang, J. J.
Carpenter, C. H.
Castor, A. H.
Chang, S. H.
Cheatham, T. P., Jr.
Chin, B. M.
Clark, R. G.
Clark, W. A., Jr.
Conley, J. W.
Cooper, Gerald
Craig, E. J.
Craig, J. W.
Cross, K. B.
Davenport, W. B., Jr.
Dennis, J. B.
Dickson, A. W.
Dieterich, E. J.
Dolansky, L. O.
Dorr, R. E.
Drouilhet, P. R., Jr.
Elias, Peter
Elkind, J. I.
Ellsworth, J. L.
Epstein, M. A.
Essigmann, M. W.
Fallows, E. M.
Fano, R. M., Dr.
Farley, B. G., Dr.
Fellows, G. E.
Fertig, Kenneth
Feuerstein, Erwin
Forbes, R. E.
Forgie, J. W.
Frank, W. I.
Fullerton, A. L., Jr.
Garber, H. N.
Gesteland, R. C.
Goldberg, W. P.
Goldman, E. M.
Goldstein, M. H., Jr.
Gottschalk, W. M.
Goulder, M. E.
Graham, J. W.
Granlund, John, Dr.
Green, P. E., Jr.
Gronroos, Eino
Gruber, Sheldon
Hall, W. M.
Hanopol, Louis
Harrington, J. V.
Heaviside, M. G.
Hedlund, D. A.
Hervey, J. P.
Hobbs, C. F.
Hofstetter, E. M.
Holt, R. B.
Huffman, D. A.
Johnson, E. P., Jr.
Keller, E. W., Prof.
Kessel, Benjamin
Kilroy, H. P.
Klein, J. A.
Krulee, R. L.
Lamere, B. E.
Lechner, R. J., Lt.
Levenstein, Harold
Lippincott, Southard
Loewenthal, Morton
Maglathlin, R. N.
McGowan, E. A.
McLaughlin, W. C.
Mercer, Wm. R.
Merfeld, Eugene
Merriam, C. W., III
Middleton, David
Mooers, C. N.
Mudgett, W. L.
Mullen, J. A.
Nagy, Ferenc, Jr.
Neumann, H. D.
Nuchols, A. J.
Ockene, N. J.
Oettinger, A. G.
Olsson, E. A.
O'Meil, S. J.
Parke, N. G., III
Perez, A. A.
Platt, H. J.
Pontecorvo, P. J.
Prager, R. H.
Radford, W. H.
Rhyins, R. W.
Richards, P. I., Dr.
Rochefort, J. S.
Rubin, M. D.
St. Jean, L. E.
Samson, E. W.
Sanders, R. C., Jr.
Schoen, E. T.
Scripture, F. R.
Seifert, W. W.
Semon, W. L.
Shekel, Jacob
Sherman, Herbert
Shnitzler, R. K.
Silverman, R. A.
Simons, J. C., Jr.
Skolnik, M. I.
Smith, T. B.
Spencer, R. C.
Stein, Seymour
Stone, M. L.
Stone, Noel
Strum, P. D.
Stuart, R. W.
Stutt, C. A.
Sussman, S. M.
Swing, R. E.
Thayer, A. R.
Tierney, Joseph
Tyler, V. M., II
Van Meter, David
Wargo, J. J.
Watson, J. K.
Webster, F. A.
Wellman, Bertram
White, G. M.
Wildes, K. L.
Williams, D. G.
Wolniewicz, R. S.
Woods, R. E.
Young, F. M.
Youngblood, W. A.
Zieman, H. E.

Buffalo-Niagara

Barnes, J. A.
Bondy, M. F.
Davis, J. E.
Earshen, J. J.
Fisher, F. M.
Freeman, R. W
Fryer, W. D.
Hardy, L. H.
Harvey, N. L.
Holmes, W. S.
Joseph, R. D.
Kay, R. H.
Kell, R. E.
Lawton, J. G.
Lerner, Theodore
Lowitz, Gabriel
Markes, R. T.
Morrison, H. V.
Novo, R. B.
Nuckolls, R. G.
Paige, Arnold
Shub, Lionel
Spoon, G. E.
Walbesser, W. J.
Weatherbee, W. H.

Connecticut Valley

Agule, G. J.
Allen, T. F.
Bogan, S. C.
Dickon, A. J.
Geiser, D. T.
Hall, B. A.
Hallenbeck, R. A.
Hazard, W. R.
Herrera, J. C.
Hildebrandt, W. J.
Holloway, H. R.
Lerner, I. S.
Loda, C. J.
Loeb, J. M.
Mapes, T. J.
Marchand, Nathan
McCoy, D. S.
Nelson, R. A.
Nuttall, A. H.
O'Keefe, E. W.
Ordung, P. F.
Osbahr, B. F.
Radom, Robert
Rettenmeyer, P. F.
Schultheiss, P. M.
Smith-Vaniz, W. R., Jr.
Stout, T. M.
Thompson, G. S.
Wesley, C. A.
Wogrin, C. A.

Elmira-Corning

Lempert, Joseph

Ithaca

Bolgiano, Ralph, Jr.
Burrows, C. R., Dr.
Charton, P. W.
Cohen, R. S.
Harmuth, H. F.
Harris, S. M.
Hausz, Walter
Horton, W. H.
Ingalls, C. E.
Jacomini, O. J.
Jankowski, Herman
Kantor, Gideon
Levinthal, J. G.
McGaughan, H. S.
Nichols, Benjamin
Perry, F. M.
Smith, F. G., Jr.

Rochester

Brown, G. A.
Chesna, John
Ferry, R. N.
Forbes, E. J.
Morse, J. E.

Rome-Utica

Blum, Harry
Davis, Harry
Grisetti, R. S.
Herson, J. L.
Jones, E. D., Jr.
Kunze, A. A.
Mackey, C. L.
Morelli, Michael
Moskowitz, Fred
Walker, R. K.

Schenectady

Abraham, L. G., Jr.
Beers, R. F.
Chernoch, Joseph
Fick, C. G.
Frum, Alexander
Hamren, S. D.
Jakowatz, C. V.
Pester, R. F.
Samsel, R. W.
Shuey, R. L.
Zawilinski, J. L.

Syracuse

Allen, J. E.
Applebaum, S. P.
Arsem, A. D.
Bartley, W. P.
Chapman, J. K.
Chynoweth, W. R.
Forney, J. F.
Fye, R. P.
Garber, S. M., Jr.
Goldman, Stanford
Hellerman, Herbert
Hu, Ming-kuei
Kahal, Robert
Millman, G. H.
Plusc, Igor
Rausch, R. H.
Samulon, H. A.

chaffner, J. S.
Travis, J. B., Jr.
Terner, C. P.
Wilcox, J. F.
Woodward, G. F.

Region II

Long Island

Abe, Teruo
Adams, B. L.
Adolph, A. R.
Agree, Irvin
Alter, Joseph
Arnell, E. B.
Atwood, J. B.
Averett, S. S.
Bailey, W. F.
Bellare, David
Bendett, R. M.
Bicknese, F. H.
Bodow, Joseph
Brandinger, J. J.
Briskin, H. B.
Brockner, C. E.
Browder, J. E.
Brownman, H. L.
Bruce, G. B.
Byron, Melville
Chadwick, J. H., Jr.
Crist, P. W.
Crosby, M. G.
Crowder, H. A.
Crowhurst, N. H.
D'Amico, S. P.
Dean, M. A.
Dettinger, David
Di Franco, J. V.
Di Toro, M. J., Dr.
Dong, H. L. H.
Dressler, Robert
Egan, W. G.
Eichel, J. H.
Evans, James
Fine, A. M.
First, Douglas
Firth, L. G., Jr.
Fogel, L. J.
Fox, Jerome
Frantz, W. P.
Freeman, E. E., Capt.
Freeman, Herbert
Friedman, E. D.
Gauger, R. H.
Giulianelli, R. M.
Goodstein, Julian
Gretz, R. W.
Gurevics, David
Hansel, P. G.
Hagan, T. G.
Hauser, A. A., Jr.
Herrburger, E. C.
Hershler, Abe
Hevesi, C. F.
Hinchey, R. L.
Hollywood, J. M.
Immarco, A. E.
Jagoe, R. W.
Johl, M. J.
Kabrisky, M. J.
Kahn, L. R.
Katz, S. H.
Kawaller, Jordan
Klein, Herman
Klein, M. M.
Kornbulth, A. E.
Kratsch, P. H.
Landauer, W. E.
Lebenbaum, M. T.
Lewinstein, Marcus
Loughlin, B. D.
Lowman, R. V.
Lustig, H. E.
Marston, R. S.
Max, S. M.
McPherson, D. L., III
Mitchell, W. M.
Mundo, C. J., Jr.
Nelson, R. T.
Norman, R. H.
Orazio, A. F.
Palmer, Winslow
Paterno, P. M.
Paull, Marvin
Peterson, H. O.
Potter, J. T.
Press, Martin
Proskauer, Richard
Rehberg, C. F.
Reitlinger, Adolf
Relis, M. J.
Richman, Donald
Ruvin, A. E.
Saks, H. L.
Schreitmueller, R. F.
Schwartz, Bernard
Schwartz, L. S.
Shostak, Michael
Simon, J. M.
Smith, H. H.
Smith, P. G.
Sporn, S. R.
Stark, E. W.
Stateman, M. F.
Stiefel, R. C.
Stuart, A. F.
Taylor, J. M.
Turner, Edward
Urbanik, J. G.
Vogel, Erwin
Volpe, C. D.
Wang, C. C.
Wappler, A. C.
Weston, Simeon
Wheeler, H. A.
White, W. D.
Winzemer, A. M.
Yang, Chia-chih
Young, V. J.
Zellers, J. E., Jr.

New York

Abell, G. R., Jr.
Alexander, B.
Alexander, Henry
Amarel, Saul
Athey, S. W.
Aubrey, Gerald
Bachman, G. F.
Banner, Leonard
Barron, Louis
Baum, M. C.
Baum, R. F.
Baumann, R. H.
Bemis, P. S.
Bennett, H. S.
Bergen, A. R.
Berger, F. B.
Bergman, Samuel
Bernstein, M. H.
Bernstein, Robert
Bifano, V. J.
Blatt, Howard
Blaustein, P. H.
Bloom, F. J.
Bond, F. E.
Brillouin, L. N.
Brown, Adamant
Brown, F. L.
Brown, P. M.
Bruns, W. H.
Burkhard, H. F.
Cameron, T. G.
Capon, Jack
Carter, W. S., Jr.
Cassidy, J. J.
Chako, Nicholas
Chang, S. S. L.
Chanzit, Lawrence
Chen, Tsung-shan
Claassen, F. S.
Clarke, E. L.
Clarke, K. K.
Cooper, F. S.
Coughlin, J. J.
Cravis, Howard
Dellenbaugh, F. S., III
Dorrell, C. E.
Duffendack, O. S.
Edling, E. A.
Ehrman, Leonard
Eisele, W. A.
Epstein, Milton
Exselsen, R. M.
Feldman, E. F.
Flores, Ivan
Fontaine, A. B.
Frank, R. L.
Freed, Arthur
Freiman, M. J.
Freudenberg, Boris
Friedland, Bernard
Froehlich, Cecilie
Gamertsfelder, G. R.
Gatzke, F. W.
Gauvenet, A. J.
Geduldig, D. S.
Gersch, W. M.
Gery, S. W.
Gilman, G. W.
Glantz, L. M.
Glowalla, John
Golay, M. J. E.
Goldberg, Bernard
Golden, R. K.
Goldman, Simon
Goldmuntz, L. A.
Goldsmith, A. N., Dr.
Goodall, W. M.
Goronin, Abraham
Green, J. R.
Greibach, E. H.
Greig, J. H.
Gronner, Alfred
Gross, Stanley
Gruenberg, E. L.
Haibt, L. H.
Hall, Wm. W., Jr., Capt.
Halpern, H. M.
Hamer, Howard
Harris, Lawrence
Hiltner, E. B.
Hogan, J. V. L.
Honig, W. M.
Jahn, D. M.
Jaworski, A. M.
Joyce, M. V., Jr.
Julie, Loebe
Kahn, Alfred
Kahn, W. K.
Korn, S. J.
Kovarik, Henry
Krauer, Otto
Kuch, B. L.
Laemmel, A. E.
Lampert, Leon
Lederman, Albert
Lee, Chi-yuan
Lerner, David
Liebman, P. M.
Lippel, Bernard
Mack, Alfred
Madsen, J. F.
Maedel, G. F.
Malchman, F. L.
Marsh, H. S.
Mastrandrea, R. J.
McClean, J. F
Mealy, G. H.
Mertz, Pierre
Meslener, G. J.
Miller, S. E.
Mitchell, Doren
Mongan, A. J.
Mountain, J. D.
Nisenoff, Norman
Nordyke, H. W., Jr.
O'Neill, L. H.
O'Sullivan, T. J.
Orefice, G. T.
Parr, A. F. W.
Pearson, K. A.
Perlman, Sol
Peterson, W. W.
Petrides, Thracy
Pickholtz, R. L.
Quinn, J. C., Jr.
Rask, O. N.
Reader, Malcolm
Reynolds, Gibson
Rice, S. O.
Rochester, Nathaniel
Rojak, F. A.
Rosen, Arnold
Rosenberg, Paul
Ross, A. H.
Roth, Jesse
Roth, S. S.
Rothstein, Jerome
Rowe, H. E.
Rowley, B. G. H.
Russell, W. J., Jr.
Rutz, R. F.
Sabto-agami, J. C.
Samuel, A. L., Dr.
Schillinger, A. G.
Schorr-kon, J. J.
Schwarz, R. J.
Sears, C. B.
Selender, Harold
Sher, F. D.
Sheridan, T. R.
Sherry, L. I.
Shivack, I. M.
Sklansky, Jack
Smith, E. J.
Solomonoff, R. J.
Speller, J. B.
Spencer, G. D.
Stern, Erick
Stockman, W. E.
Strasser, Stella
Tabor, J. H., Dr.
Taylor, W. H., Jr.
Truxal, J. G.
Tuckerman, L. S.
Unger, S. H.
Vaughan, J. A.
Wallace, Marcel
Walston, C. E.
Weber, Ernst, Prof.
Weber, L. A.
Whitenack, R. M.
Whittle, R. L.
Wilson, I. G.
Winson, Philip
Wise, R. J.
Wolensky, William
Wolotkin, Aaron
Wynne, W. M., Comdr.
Zadeh, L. A.
Zaslavsky, Sam
Zemanian, A. H.

Northern New Jersey

Aaron, M. R.
Anderson, G. M.
Bagno, S. M.
Bailey, R. S.
Baldwin, G. L.
Bennett, W. R.
Black, H. S.
Brogan, J. M.
Bucher, F. X.
Burford, T. M.
Busignies, H.
Cahill, W. J.
Caretta, P. M.
Clavier, A. G.
Compton, H. B.
Cotellessa, R. F.
Crofts, G. B.
David, E. E., Jr.
Davis, R. C.
Dayem, A. H.
De Graaf, D. A.
De Rosa, L. A.
De Vany, B. V.
Dickey, D. W.
Edison, T. M.
Felch, E. P.
Fischer, L. G.
Goeller, L. F., Jr.
Gordon, M. J.
Graham, R. E.
Grandmont, P. E.
Groce, J. C.
Gross, F. J.
Grossman, A. J.
Grower, B. B.
Guenther, Richard
Hartley, R. V. L.
Hays, F. R.
Heiser, W. H.
Hibbert, J. J.
Hoberman, M. J.
Hulst, G. D.
Jensen, A. G.
Kamentsky, L. A.
Kang, Chi Lung
Kreer, J. G., Jr.
Kretzmer, E. R.
Kups, E. F.
Lewis, R. F.
Lundburg, F. J.
Lupo, F. J.
Martin, J. F.
Martin, S. J.
Mathews, M. V.
McMillan, Brockway
Meacham, L. A.
Meier, W. L.
Mock, W. C., Jr.
Montcalm, S. R.
Moore, E. F.
Murdock, B. B.
Nestor, George
Perry, A. D., Jr.
Rea, W. T.
Rizzo, J. E.
Rogoff, Mortimer
Ronald, T. T.
Rubino, J. W
Saloom, J. A.
Sauber, J. W.

Schott, L. O.
Schwartz, E. I.
Seckler, H. N.
Shangraw, C. C.
Shannon, C. E.
Sherman, Seymour
Siekanowicz, W. W.
Slepian, David
Sudduth, W. B.
Sullivan, H. J.
Sweeney, W. R.
Terry, C. B.
Thompson, T. H.
Treuhaft, M. A.
Turkheimer, P. M.
Van Tassel, E. K.
West, S. S.
Westbrook, E. P.
Wickliffe, P. R., Jr.
Wing, Omar
Wirth, C. H.
Zayac, F. R.

Princeton

Bellew, F. N.
Cadden, W. J.
Ceres, G. V.
Clement, P. R.
Fich, Sylvan
Gilchrist, Bruce
Lindsay, F. A.
Mather, N. W.
Shomer, J. E.
Strother, J. A.
Surber, W. H., Jr.
Sziklai, G. C.
Wong, S. Y.

Region III

Atlanta

Corriher, H. A., Jr.
Dasher, B. J.
Emmert, R. A.
Finn, D. L.
Flynt, E. R.
Mauldin, H. W., Jr.
Roberts, T. E., Jr.
Schultz, F. V.
Walker, E. M.
Widerquist, V. R.

Baltimore

Basil, I. T.
Blasbalg, Herman
Broady, S. N.
Browdin, M. E.
Cichanowicz, H. J.
Csepely, J. A.
Esterson, G. L.
Gilbert, G. B., Jr.
Glaser, E. M.
Goetz, L. P.
Gott, Euyen
Herwald, S. W.
Huggins, W. H.
Isley, C. T., Jr.
Jenkins, Jobe
Jones, L. G. F.
Kovasznay, L. S. G.
Mester, J. C.
Mott, O. B.
Pierson, C. D., Jr.
Pipkin, J. E.
Pullen, K. A., Jr.
Raboy, Bernard
Rambo, S. I.
Rosenblum, I. I.
Schultz, R. L.
Terashi, Taro
Thomas, J. F.
Wedge, T. E.
Worrell, E. A.

Central Florida

Anderson, G. F.
Downs, J. W.
Friedland, M. S.
Hines, W. S.
Koning, R. E.
Lahn, F. C.
Lloyd, L. H.
Neidert, J. H.
Painter, Parker, Jr.
Rogers, C. L., Jr.

Huntsville

Clements, C. C.
Greenwood, T. L.
Schumann, Fred

Miami

Kush, L. J.
Wise, D. O., Jr.

North Carolina-Virginia

George, R. T., Jr.
Gremer, C. E., Lt.
Harris, O. R.
Hastings, C. E.
Passera, A. L.
Pitsenberger, J. W.

Philadelphia

Annett, M. E.
Bachofer, H. L.
Balakrishnan, A. V.
Bargellini, I. P. L., Dr.
Beaumariage, D. C.
Benner, R. H., Dr., II
Bensky, L. S.
Berkowitz, R. S.
Bose, A. G.
Bradford, C. E.
Buchanan, J. P.
Bucher, T. T. N.
Byers, A. C.
Cecala, J. A.
Charton, Seymour
Cohen, Irving
Cramer, B. G.
Deutsch, Joseph
Drenick, Rudolf
Ehrich, W. G.
Eichert, E. S.
Emslie, N. M.
Foley, G. M.
Friend, A. W., Dr.
Gans, J. S.
Gardner, R. K.
Garrett, L. B., Jr.
Gaynor, Emil
Gollub, Raphael
Gottschalk, J. M.
Greene, A. H.
Greenfield, Alexander
Halev, Irvin
Hobbs, L. C.
Holshouser, J. R., Jr.
Honda, Hajime
Hopper, G. M., Dr.
Hurford, W. L.
Kassler, Raymond
Katz, Abraham
Katz, E. S.
Kerfoot, B. P., Jr.
Ku, Yu H.
Lazinski, R. H.
Linden, D. A.
Markarian, Hagop
Maron, Irving
Mathes, R. E.
Meredith, B. E.
Merrihew, H. W.
Mertens, L. E.
Million, J. W., Jr.
Morton, W. B., Jr.
Mrazek, D. A.
Noel, R. Y. M.
Patterson, G. W.
Price, Robert
Ringer, H. N.
Roberts, F. R.
Roberts, M. J.
Rogers, R. F.
Roseman, H. M.
Rosen, George
Rosenzweig, G. E.
Ryan, V. F., Jr.
Sachs, H. L.
Schroeder, R. P.
Schulz, R. B.
Showers, R. M.
Sorkin, C. S.
Steinberg, B. D.
Tou, Julius, Dr.
Urkowitz, Harry
Weisbecker, J. A.
Weiss, Eric
Williams, A. J., Jr.
Wolin, Samuel
Woll, H. J.
Yamada, Hisao
Yang, Tsute

Washington, D. C.

Alde, R. O.
Alderson, W. S.
Ballard, A. H.
Bates, J. F.
Bauer, P. S., Cdr.
Baxter, J. L.
Beck, H. M.
Blackburn, C. A.
Blair, C. R.
Calhoon, T. G.
Campanella, S. J.
Cook, F. W.
Cooper, H. W.
Cowan, Bryan, Lt. Col.
Donnell, W. F.
Elbourn, R. D.
Fine, Harry
Finn, P. L.
Finney, W. J.
Fleming, J. J.
Friedberg, I. S.
George, S. F.
Gerig, J. S.
Geselowitz, D. B.
Gleason, R. F.
Godsey, W. J.
Goldberg, Harold
Grimstad, E. J.
Grisamore, N. T.
Hafer, F. L.
Hawthorne, G. B., Jr.
Haydon, G. W.
Headrick, J. M.
Hedge, L. B.
Hepler, D. S.
Heyliger, G. E.
Hogan, D. L.
Holman, J. G.
Holmes, C. H.
Horton, B. M.
James, W. G.
Karr, P. R.
Katzin, Martin
King, A. M.
King, W. P.
Kirsch, R. A.
Kirshner, J. M.
Klein, M. H.
Kohler, H. W.
Kriz, Joseph
Kuck, J. H.
Kullback, Solomon
La Pointe, J. C.
Larson, R. C.
Leiner, A. L.
Levy, J. E.
Lieberman, Gilbert
Lord, J. B.
Lun, M. J.
McCracken, L. G., Jr.
McGinnis, C. E.
McClurg, G. H.
Melton, B. S.
Moore, C. G., Jr.
Neumann, A. J.
Ould, R. S.
Page, C. H., Dr.
Page, R. M.
Parr, E. D.
Peterson, H. L.
Phillips, M. L., Mrs.
Polak, Henri
Potter, R. W.
Powell, R. M.
Reed, S. F.
Rotkin, Israel
Ryan, A. H.
Salerno, James
Schwartz, R. J.
Scott, R. M.
Scott, S. R.
Scott, W. H., Jr.
Sears, J. F.
Sherertz, P. C.
Smith, B. D., Jr.
Smith, E. L., Jr.
Spriggs, J. O.
Stastny, G. F.
Stockebrand, T. C.
Sugar, G. R.
Summers, C. R.
Talkin, A. I.
Tieman, C. R.
Thompson, R. T., Jr.
Wald, Bruce
Waldschmitt, J. A.
Weiland, F. A.
Wilcox, R. H.
Willard, J. M.
Wright, W. W.
Youden, W. W.
Young, J. W., Jr.
Zakowski, L. F.
Zirm, R. R.

Region IV

Akron

Bushnell, R. H.
Diamantides, N. D
Flowers, H. L.
Gaul, E. R.
Kult, M. L.
Pressel, P. I.
Richman, M. A.

Cincinnati

Beyer, J. P.
Doerr, W. H.
Edwards, R. L., Jr.
Grosch, H. R. J., Dr.
Willsey, R. H.

Cleveland

Baerwald, H. G.
Ehrman, J. R.
Hare, V. C. M., Jr.
Hitt, J. J.
King, C. F
Kres, A. J.
Laden, H. N., Lcdr.
Phillips, W. E., Jr.
Pulaski, M. E.
Wickenden, H. R.

Columbus

Chope, H. R.
Clifton, J. K.
Dawirs, H. N.
Warren, C. E.
Weimer, F. C.

Detroit

Batten, H. W.
Bernhard, H. A.
Book, Everard
Carr, J. W., III
Cutrona, L. J.
Deutsch, Menachem
Farris, H. W.
Garner, H. L.
Gilbert, E. O.
Gilbert, E. G.
Goodrich, R. H.
Gottesman, N. H.
Hok, Gunnar
Jacobs, M. A
Lindsay, W. J.
Longerich, E. P.
Macnee, A. B.
McPherson, R. R.
Morin, D. C., Jr.
Nakagawa, Noriyuki
Otterman, Joseph
Piper, C. A.
Rauch, L. L.
Reiher, H. F.
Rupcich, J. N.
Schoderbek, J. J.
Scott, N. R.
Sharpe, C. B.
Stewart, J. L.
Szajna, E. F.
Tu, Ju. C.

Emporia

Baker, W. L.
Byers, H. K.
Golla, E. F.
Harvey, H. B.
Higdon, R. V.
Key, C. L., Jr.
Knausenberger, G. E.
Lawther, J. M.
Lopez, A. F.
Myers, S. J.

Pittsburgh

Adams, C. W.
Alexander, F. C., Jr.

ight, R. L.
ırry, T. F.
ean, W. C.
rd, D. J.
ınnon, G. F., Jr.
ace, J. N.
ıet, F. R., Mrs.
ıtchison, T. C.
ollandbeck, R. F.
otzbaugh, G. A.
arlowe, E. W.
arshall, B. O., Jr.
Donnell, T. J.
hatz, E. R.
ın Nice, R. I.
oodford, J. B., Jr.

oledo

ıjii, Takashi

'illiamsport

ebb, H. E.

Region V

edar Rapids

ıbillus, John
owenberg, E. C.
over, H. A.
ilson, D. R.

ıicago

ierer, J. C.
ıderson, J. M.
all, R. B.
eam, R. E.
erry, R. F.
elawa, F. R.
orrowman, J. H.
auer, H. H.
ırlson, G. R.
ırter, Robert
avier, P. A.
ohn, G. I.
ohn, Jona
ooney, J. J.
ruz, W. S.
restone, W. L.
alejs, Janis
erlach, A. A.
ımpel, D. J.
oslovich, S. J.
upert, J. J.
ırvis, K. W.
nes, R. W.
arow, K. A.
lem, R. F.
ırson, R. W.
ewis, H. A.
ndholm, C. R.
ong, J. F., Jr.
agnuski, Henry
ansfield, Ralph
essinger, H. P.
oon, R. J.
orrison, Peter
lson, W. R.
zker, Tarik
ıradise, M. E.
ve, H. C.
ichards, H. F.
uina, Jack
ıltzberg, Bernard
ırahan, B. L.
ıapin, Theodore, Jr.
oma, Raymond

Sullivan, J. M., Jr.
Tyksinski, S. P.
Warnke, G. F.
Webb, H. D.
Weissman, R. M.
Yuen, P. C.

Dayton

Barker, W. J.
Bordewisch, J. F.
Brown, G. T., Jr.
Goldman, C. C.
Gould, G. P.
Haneman, V. S., Jr.
Kott, W. O.
Lockwood, G. C.
McLennan, M. A.
Piety, E. W.
Thompson, C. W. N.
Waugh, R. E.

Des Moines-Ames

Leverington, R. D.

Fort Wayne

Chambers, W. A.
Johnson, D. L.
Weidner, E. J.

Indianapolis

Clark, J. R.
Cooper, G. R.
Fenimore, G. E.
Goldstein, S. J., Jr.
McCune, D. A.
Pitt, P. E.

Milwaukee

Asmuth, J. L.
Halijak, C. A.
Herzog, Will
Rideout, V. C.
Scheibe, E. H.
Theiss, C. M.

Omaha-Lincoln

Chamberlain, I. J.
Wycoff, K. H.

Twin Cities

Bashara, N. M.
Bratschi, R. W.
Cohen, A. A.
Essler, W. O.
Featherstone, R. P.
Gergen, J. L.
Hardenbergh, G. A.
Harris, L. A.
Kelsey, J. R.
Koschmann, A. H.
Lode, Tenny
Ludwig, J. T.
Markusen, D. L.
Morell, C. S.
Rowland, C. A., Jr.
Schmitt, O. H.
Youngquist, R. J.

Region VI

Dallas-Fort Worth

Brachman, M. K.
Brust, M. F.
Dodd, J. M.
Dolan, B. A., Lt.
Fletcher, C. H.
Gudzin, M. G.
Harmon, F. I.
Jones, H. J.
Leming, T. L.
Mitchell, W. R.
Muerle, J. L.
Mut, S. C.
Wadel, L. B.
Weedfall, W. W.
Wright, T. A.

Denver

Carlin, P. W.
Cottony, H. V.
Daniels, W. H.
Fulton, F. F., Jr.
Slutz, R. J.

El Paso

Carbine, I. L.
Kidwell, R. P.
Kutrumanes, C. P.
Lovitt, S. A.
Mickleburgh, W .C., Pvt.
Miller, W. E., Jr.
Muehlner, J. W.
Pyle, C. A.
Sherburne, R. K.
Smith, K. E.
Thompson, D. I.
Wagner, Nathan
Yarter, B. P.

Houston

Ball, J. D.
Chapman, D. C.
Easterling, M. F.
Harbaugh, A. W.
Harker, H. J.
Keating, L. M.
Melloh, A. W.
Pierson, A. L., III
Rust, W. M., Jr.
Tanguy, D. R.
Weber, M. E.
Wischmeyer, C. R.

Kansas City

Miller, H. G.
Pulliam, C. S.

New Orleans

Dowe, R. J.
Gordon, C. K.
McLean, L. V.

Oklahoma City

Challenner, A. P.
Daniel, D. B.
Puckett, T. H.

St. Louis

Fordyce, S. W.
Furfine, A. L.
Hirsch, O. C.
Keiser, B. E.
Little, G. R.
Mohrman, R. F.
Palmer, J. A.
Roberts, E. L.
Tedeschi, Anthony
Van Bladel, J. G.
Winter, D. F.

San Antonio

Douglas, J. H.
Hoffman, A. A. J.
Levin, M. J., Pvt.
Ziemer, D. R.

Tulsa

Day, C. E.
Kammerzell, C. E.
Moxley, S. D., Jr.
Piety, R. G.
Rice, R. B.
Scott, C. B.
Silverman, Daniel

Region VII

Albuquerque-Los Alamos

Banks, T. G., Jr.
Basore, B. L.
Bidwell, H. H.
Brown, W. E., Jr.
Fursa, Alex
Moore, R. K.
Skinner, L. V.
Williams, C. S., Jr.

Inyokern

Zilmer, D. E.

Los Angeles

Abramson, N. M.
Ackerlind, Erik
Adrian, D. J.
Albrecht, Albert
Allen, D. H.
Andrews, L. A.
Antell, Stanley
Apa, F. E.
Asawa, C. K.
Ashby, R. M.
Ausbourne, R. K.
Avrin, J. S.
Babcock, D. F.
Barnes, J. L.
Beck, Leonard
Bedrosian, Edward
Bell, N. W.
Begovich, N. A., Dr.
Bird, R. N.
Bower, J. L.
Boyd, W. L.
Brady, F. H.
Braun, E. L.
Brennan, L. E.
Briggs, J. G.
Brown, C. S.
Burkart, E. H.
Cain, G. H., Jr.
Campbell, R. A.
Carlson, C. O.
Chu, Henry
Colander, R. E.
Culver, W. H.
Cummings, C. I.
Curl, G. W.
Davis, F. W.
Davis, Harold
Davis, J. S.
De Lano, R. H.
Dethlefsen, D. G.
Deutsch, Ralph
Diemer, F. P.
Downes, Lloyd
Drake, W. A.
Duncan, D. B.
Du Waldt, B. J.
Edelsohn, C. R.
Endsley, G. T.
Epstein, R. A.
Escher, P. H.
Fatton, G. A.
Fisher, R. L.
Fishman, Max
Foxman, Eugene
Francis, J. P.
Frankel, S. P.
Fuller, R. H.
Garber, L. F., Lt.
Gardner, L. B., Dr.
Gates, C. R.
Gates, H. P., Jr.
Gerardi, F. R.
Gilchriest, C. E.
Gola, A. S.
Green, D. J.
Greenbaum, Marvin
Gross, William
Hadden, F. A.
Hance, H. V., Dr.
Hannum, A. J.
Hare, G. H.
Hayes, W. T.
Heilfron, Jack
Hodson, W. G.
Holly, C. M.
Howard, S. L.
Inouye, G. T.
Jacobs, J. E.
Jacobson, R. E., Jr.
Joerger, J. C.
Kelly, D. H.
Klein, W. J. J.
Klestadt, Bernard
Knopoff, Leon
Krill, C. K.
Lader, L. J.
Lambert, J. M.
Lambert, J. D.
Landman, R. M.
Larse, G. L.
Lehan, F. W.
Lephakis, A. J.
Levinson, R. M.
Lew, Clinton
Louie, William
Low, Henry
Lyons, L. H.
MacIntyre, R. M.
Magnuson, V. P.
Maki, G. J.
Mathews, W. E.
McCormick, G. F.
McFarlane, M. D.
McNabb, J. W.
McRuer, D. T.
McVey, B. D.
Molloy, C. T.
Moore, J. R.
Moreno, C. A.
Muchmore, R. B.
Munushian, Jack
Noland, A. R.
Nunn, W. M., Jr.
Nussbaum, O. N.
Parker, N. F.
Pfeffer, Irwin
Pierson, J. E.
Politi, E. Y.
Proud, W. H.
Rawlins, R. E.
Rechtin, Eberhardt

Reedy, P. H.
Reilly, Michael
Roberson, R. E.
Rosenstein, A. B.
Rumer, W. I.
Rutkin, B. B.
Rypinsk, C. A., Jr.
Salzer, J. M.
Samuelson, H. R.
Schalk, Norbert
Schindler, Mark
Schreiber, W. F.
Seltzer, L. J.
Sensiper, Samuel
Silva, L. M.
Slogar, J. W.
Snyder, W. A.
Starr, A. R.
Stimpson, L. D., Jr.
Stoehr, W. F.
Stoltz, J. R.
Stoltz, P. G.
Taber, J. E.
Tatum, F. A.
Thomsen, R. K.
Toeppe, W. J., Jr.
Trautman, D. L.
Van Horne, T. B.
Votava, Yaro
Waddell, B. L.
Wagner, D. W.
Walp, R. M.
Walquist, R. L.
Wanlass, S. D.
Ware, W. H.
Warner, S. D.
Wedel, J. J., Jr.
Weidman, J. S.
West, G. P.
Westlake, P. R.
Whiteley, T. B.
Whitford, R. K.
Wiggins, E. T.
Williams, R. D.
Wohl, Jack
Wood, B. C.
Wright, P. B.
Young, C. W.
Young, G. O.
Zumba, C. F.

Phoenix

Albright, A. R. G.
Bard, W. E.
Brooks, H. B.
Hammond, J. R.
Morgan, H. L.
Morrison, Fred
Noon, J. R.
Perper, Lloyd
Ross, J. M.
Thomas, S. M.
Winkler, Stanley

Portland

Donoghue, J. J.
Goldberg, P. A.
Ropiequet, R. L.
Strain, D. C.

Salt Lake City

Davidson, R. A.
Marsden, R. S., Jr.

San Diego

Austin, R. W.
Caspers, J. W.
Gottwald, C. H.
Hardy, F. J.
Kimball, J. L.
Lenihan, Jeremiah, Lt.
Loeb, Marvin
Meckfessel, E. F.
Ogram, R. L.
Schreiber, O. W.
Simmons, J. M., Jr.
Torres, J. F.
Wade, Ernest
Wallis, W. R., Cdr.
Werner, H. C.
Zable, W. J.

San Francisco

Allen, T. L., Jr.
Allison, J. E.
Amara, R. C.
Baker, R. H.
Bandtel, K. C.
Barnard, G. A., III
Blachman, N. M.
Blanchard, H. P.
Borghi, R. P.
Budd, W. E.
Chodorow, Marvin
Christie, J. W.
Church, Randolph
Clemens, G. W., Jr.
Cornwell, R. C.
Davies, L. E.
Dell, H. R.
Demuth, H. B.
Downie, W. A.
Elspas, Bernard
Goldberg, Jacob
Gong, H. J.
Granger, J. V. N.
Haanstra, J. W.
Halina, J. W. O.
Harman, W. W.
Haskell, H. B.
Hill, C. M.
Honey, J. F.
Jaynes, E. T.
Jones, E. D.
Karp, A. L.
Kautz, W. H.
Keitel, G. H.
LaBree, C. T.
Leifer, Meyer
Ludovici, B. F.
Matthaei, G. L.
Meader, H. W.
Moore, E. J.
Morton, P. L.
Nelson, E. A.
Nilsson, N. J.
Norcross, J. B.
Oliver, B. M.
Pappas, N. L.
Schrader, G. F.
Serniuk, Walter
Siegman, A. E.
Strassner, R. M.
Stubbins, W. F.
Swarm, H. M.
Taylor, John
Thomas, J. B.
Tillotson, J. H.
Tuttle, D. F., Jr.
Vreeland, R. W.
Wenger, D. B.
Whinnery, J. R.
Whitson, A. L.
Zeidler, H. M.

Seattle

Betts, A. L.
Bishop, D. J.
Drummond, W. D.
Erdman, J. P.
Good, J. L.
Linder, W. J.
Maynard, J. E.
Nutting, D. C.
Keller, N. E.
Klee, B. J.
Roe, G. F., Jr.

Hawaii

Stagner, G. H.

Region VIII Canada

Hamilton

Carnahan, C. W.
Moore, B. W. L.
Ratz, H. C.

London

Dearle, R. C.

Montreal

Berube, J. P. G.
Birman, Gerhard
Caron, J. Y.
Glegg, K. C. M.
Johnston, R. F.
Kassner, M. H.
Kelsey, E. S.
Lalonde, R. P.
Lortie, A. L.
Reeves, Rene
Richard, G. B.
Robichard, L. P. A.
Somers, Howard
Vaillancourt, R. M.

Ottawa

Dunlop, D. P.
Galbraith, R. A. H.
Glinski, George
Lazecki, Stanislaw

Toronto

Byers, H. G.
Cotterill, M. J.
Isserstedt, S. G.
Jagger, C. E.
Kates, Josef
Keeping, K. J.
Lang, G. R.
Mittra, Rajjeshwar
Newhall, E. E.
Qua, E. W.
Sinclair, George
Stewart, J. A.
Stoddart, T. W. H.
Szekely, Zoltan
Toye, J. M.
Yen, Jui Lin

Vancouver

Kersey, L. R.
Noakes, Frank
Moore, A. D.
Pirart, M. A.

Foreign

Argentina

Pinasco, S. F.

Australia

Lampard, D. G., Dr.
Swire, B. E.

Belgium

Desirant, M. C., Dr.

Bermuda

Harries, J. H. O.

Canal Zone

McLaughlin, W. J., Lt.

Ceylon

Gnanalingam, S.

Cuba

Arnaud, J. P.
Guiral, R. L.
Montes, J. V.

Denmark

Rybner, Jorgen, Prof.

England

Ellesworth, George
Harris, K. E.
Hatton, W. L.
Kinross, R. I., Maj.
Parsons, A. N., Sr.
Pollard, J. R.
Rantzen, H. B.
Ratcliffe, J. A.

France

Berline, S. D.
Blassel, P. P.
Brodin, Jean
Ferrier, P. A.
Fromageot, A. P. A.
Labin, Edouard
Labin, Emile
Simon, J. C.
Sokoloff, B. A.
Stevens, A. M.

Germany

Peters, J. F.
Walther, Alwin, Dr.

Holland

Alma, G. H. P.
Houtsmuller, J.
Koster, J. A.
Tellegen, B. D. H.

Hongkong

Fan, Sin-Pin
Hsu, Hsiung

India

Correa, W. R.
Ghose, Amalkumar
Mirchandani, I. T.
Sweet, C. M.

Israel

Hessell, Alexander
Loev, David
Mass, Jonathan
Segalov, Elli
Silberfarb, Elieser
Toth, P. A.
Zakhaim, Moshe

Italy

Biondi, Emanuele
Mazzarella, Mario
Sacerdoti, Giorgio
Silleni, S. R.
Tiberio, Ugo

Japan

Kuroiwa, Yutaka
Minozuma, Jumio
Mita, Shigeru
Nishino, Osamu
Tomono, Masami
Yasuda, Ichiji

Lebanon

Hoffman, J. D.

Mexico

Auerbach, L. F.

Netherlands

Hershberger, W. D.
Stumpers, F. L. H. M
Van Wijngaarden, Adriaan

Norway

Bostad, J.
Jenssen, Matz

Puerto Rico

Jones, T. L.

Spain

DeSobrino, Ricardo

Sweden

Aurell, C. G. P.
Elfving, A. L.
Lofgren, E. O.
Nilsson, B. N. A.
Svala, C. G.

Switzerland

Gloor, Bruno
Guanella, Gustave

Thailand

Jadisdha, Santa, Capt

Venezuela

Bartelme, R. R.
Hackett, A. H.

Overseas Military

Manfroi, N. A.
Miller, D. L., Capt.
Staples, E. B.
Tary, J. J.
Tetley, W. H., Col.
Walsh, J. P.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT TRANSACTIONS. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in india ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the PROCEEDINGS OF THE IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE TRANSACTIONS ON INFORMATION THEORY is published three times a year, in March, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. A period of approximately two months additional is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. There will be no June issue. The IRE CONVENTION RECORD is published at that time, and a bound collection of Information Theory papers to be delivered at the March 1956 IRE National Convention will be mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to Laurin G. Fischer, Federal Telecommunications Labs., 492 River Road, Nutley, N. J. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Nathan Marchand, Marchand Electronic Labs., 255 Mill Street, Byram, Conn.